Recognition of and response to deaminated bases by archaeal DNA polymerases

Tomas Takuyoshi Richardson

A thesis submitted for the Degree of Doctor of Philosophy

Institute of Cell and Molecular Bioscience

August 2011

Dedicated to Michael Gerard Richardson

Abstract

The archaea comprise one of the three domains of life and are often characterised by a propensity for physically- and geochemically-extreme environments. Under such conditions spontaneous DNA deamination events, which ordinarily occur at stochastically insignificant rates, increase in frequency to the point where specialised recognition pathways are required to maintain genomic stability. Archaeal DNA polymerases are unique in their capacity to recognise and respond to deaminated bases, such as uracil and hypoxanthine. For example, the family B DNA polymerases of archaea possess a well-characterised uracil-binding pocket, which, helps prevent replicative bypass of deaminated bases and thus proliferation of fixed mutations.

This thesis aims to elucidate additional features of the deaminated base recognition pathways of archaeal DNA polymerases. Here we present studies that concern both the family B and more enigmatic family D DNA polymerases of archaea. Methods employed for investigation of these enzymes include mobility shift assays, targeted mutagenesis, primer extension, exonuclease and uracil-DNA glycosylase assays, as well as time-resolved and steady-state fluorescence analysis. Furthermore, through genetic manipulation of *Thermococcus kodakarensis*, this work seeks to address previously unanswered questions regarding DNA replication and repair in the archaea. The *in vivo* studies of deaminated base recognition presented in chapter 5 raise intriguing questions about fundamental aspects of the molecular and cellular biology of archaea.

Acknowledgements

Firstly, I would like to wholeheartedly thank my supervisor, Professor Bernard Connolly, without whom none of this work would have been possible. Thank you to all of our collaborators, especially Tom Santangelo and John Reeve for the wonderful hospitality. My unending gratitude goes out to the rest of my family, Shizuko, Bobi and Tim, for all the help and support that they've provided over the years. I would like to thank Pauline, Stan, Brian and everyone who worked within the group at one point or another during my PhD; you have all helped to make the last four years an incredibly memorable and enjoyable time. Thanks to Ema for the excellent binding. Thanks to my partner, Katarzyna, for supporting me and putting up with so very much. However, most of all, I would like to dedicate this work to my father; a constant inspiration both in life and in death. Michael, I hope this makes you very proud.

Table of Contents

Chapter 2:

Chapter 9:

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

List of Tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Abbreviations

2AP	……	2-aminopurine
6MP	……	6-methylpurine
APS	……	Ammonium persulphate
ASW	……	Artificial seawater
BSA	……	Bovine serum albumin
CAM	……	Chloramphenicol
carb	……	Carbenicillin
dATP	……	Deoxyadenosine triphosphate
dCTP	……	Deoxycytidine triphosphate
dGTP	……	Deoxyguanosine triphosphate
DMT	……	Dimethoxytrityl
DNA	……	Deoxyribonucleic acid
DNaseI	……	DNA endonuclease I
DTT	……	Dithiothreitol
dTTP	……	Deoxythymidine triphosphate
dUTP	……	Deoxyuridine triphosphate
E. coli	……	*Escherichia coli*
EDTA	……	Ethylenediaminetetraacetic acid
Hex	……	6 - carboxy - 2',4,4',5',7,7' – hexachlorofluorescein
HPLC	……	High-pressure liquid chromatography
kan	……	Kanamycin
Kcr-Pol D	……	*Korarchaeum cryptofilum* Pol D
LB	……	Lysogeny broth
Neq	……	*Nanoarchaeum equitans*
Neq-Pol D	……	*Nanoarcheum equitans* Pol D
Nma-Pol D	……	*Nitrosopumilus maritimus* Pol D
OD	……	Optical density
Pab	……	*Pyrococcus abyssi*
Pab-Pol D	……	*Pab* family D DNA polymerase

PAGE	……	Polyacrylamide gel electrophoresis
PCR	……	Polymerase chain reaction
Pfu	……	*Pyrococcus furiosus*
Pfu-PCNA	……	*Pfu* proliferating cell nuclear antigen
Pfu-Pol B	……	*Pfu* family B DNA polymerase
Pfu-Pol D	……	*Pfu* family D DNA polymerase
Pgl	……	*Pyrococcus glycovorans*
pH	……	Potential hydrogen
Pya	……	*Pyrococcus yayanosii*
RNA	……	Ribonucleic acid
RT	……	Reverse transcription
SDS	……	Sodium dodecyl sulphate
SDS-PAGE	……	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
Taq	……	*Thermus aquaticus*
TBE	……	Tris/borate/EDTA
Tch	……	*Thermococcus chitonophagus*
TEMED	……	N,N,N',N'-Tetramethylethylenediamine
Tga	……	*Thermococcus gammatolerans*
Tgo	……	*Thermococcus gorgonarius*
Tgu	……	*Thermococcus guaymasensis*
Tgo-Pol B	……	*Tgo* family B DNA polymerase
Tko/T. kod	……	*Thermococcus kodakarensis*
UDG	……	Uracil DNA glycosylase
UV	……	Ultra violet
WT	……	Wild type

Chapter 1

Introduction

1.1: The discovery of deoxyribonucleic acid

Deoxyribonucleic acid (DNA) encodes all of the information required for living cells, as well as most acellular organisms, to function in an ordered manner. It provides the mechanism by which the vast majority of life on earth is able to propagate and maintain its biological make-up over successive generations.

While the concept of heritability in living organisms can be traced back to at least the era of Plato (Plato, c. 375 BC), the development of a workable biological model did not occur until relatively recently. Throughout history a plethora of scientists and scholars has tackled the subject, however, it was Friedrich Miescher, a Swiss physician and biologist, who first isolated DNA in 1869 (Dahm, 2008). The significance of the substance, originally termed "nuclein", was not immediately apparent and, as such, research on its properties was not particularly forthcoming. Studies, which were simultaneously performed by Gregor Mendel, helped shape our understanding of inheritance and genetics (Mendel, 1865), however, DNA research really began to progress in the early to mid-twentieth century. In 1937 Florence Bell and William Astbury performed some of the first X-ray diffraction studies on DNA, in which they established that it possesses a regular, repetitive structure (Astbury and Bell, 1938). While they were unable to solve the structure themselves, their work led directly to studies performed by Raymond Gosling, Maurice Wilkins and Rosalind Franklin. The trio identified two different forms of DNA, termed A and B, and showed that the adopted form was dependent on the water content of the sample. They produced a series of now iconic diffraction patterns, including, most-famously, "Photograph 51", which is shown in figure 1.1 (Franklin and Gosling, 1953). The diffraction pattern clearly contains a distinctive cross, indicative of the helical structure of the sample.

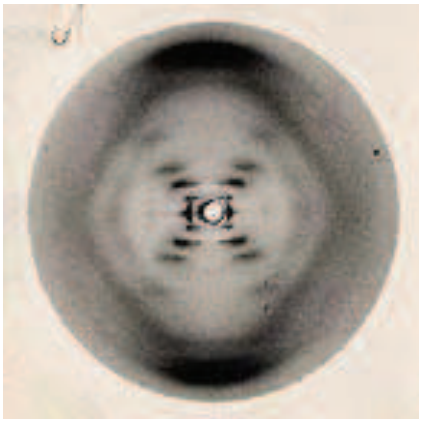


Figure 1.1: Photograph 51, produced by Raymond Gosling and Rosalind Franklin in 1952, revealing the diffraction pattern of the hydrated, B-form of DNA (Franklin and Gosling, 1953).

During this highly progressive period of DNA research Erwin Chargaff helped devise the base-pairing theory by showing that, although the relative proportions of the four bases varies between different organisms, the amount of adenine present is always equal to that of thymine and, similarly, the amount of cytosine is always equal to that of guanine (Chargaff *et al.*, 1951). Eventually, in April 1953, James Watson and Francis Crick published their seminal paper, in which they successfully proposed the correct structure of DNA (Watson and Crick, 1953).

1.2: Structure of DNA

In their undamaged native state nucleic acids exist as polymers made up of repeating units of phosphate, sugar and one of four nitrogen-containing bases; collectively known as nucleotides. The sugar-phosphate backbone of DNA, shown in figure 1.2c, is made up of alternating molecules of deoxyribose (1.2a) and phosphoric acid (1.2b), linked together by the formation of phosphodiester bonds between the 3rd (3') and 5th (5') carbon atoms of adjacent deoxyribose molecules.

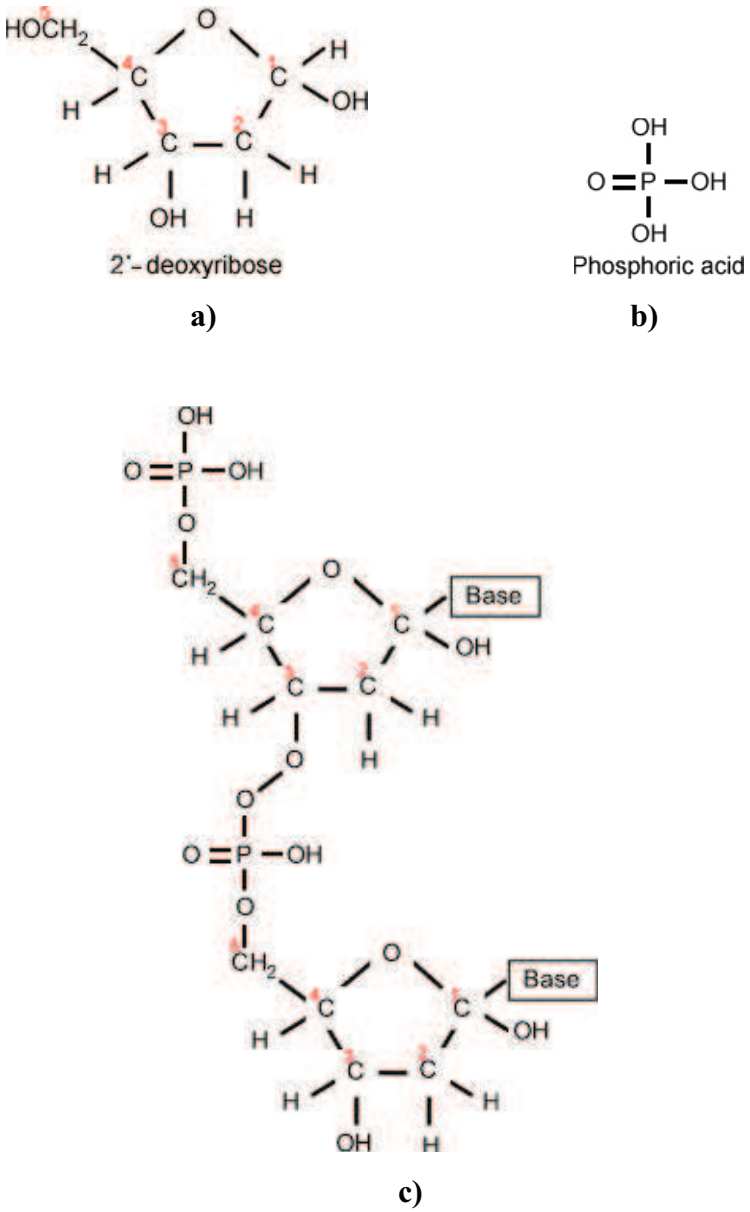


Figure 1.2: Individual components of the sugar-phosphate backbone of DNA. **a)** 2'-Deoxyribose. **b)** Phosphoric acid. **c)** Interconnected nucleotides in a single strand of DNA. Adjacent nucleotides are linked by the formation of a phosphodiester bond (adapted from http://www.dr-sanderson.org/geneticcontrol.htm).

In undamaged DNA one of four nitrogen-containing bases (adenine, guanine, cytosine or thymine) is covalently attached to the first carbon (1') of each deoxyribose molecule. Adenine and guanine are both purines, having two carbon-nitrogen ring structures. Thymine, cytosine and uracil on the other hand possess just one carbon-

nitrogen ring and are known as pyrimidines. The chemical structures of all of these bases can be seen in figure 1.3.

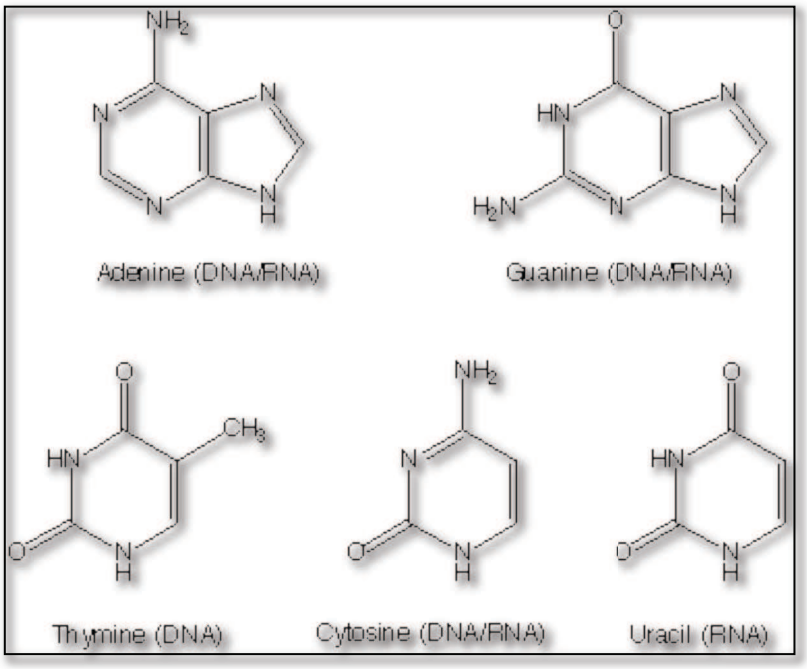


Figure 1.3: Chemical structures of bases contained in undamaged DNA and ribonucleic acid (RNA). The purines, adenine and guanine, are shown on the top row, while the pyrimidines, thymine, cytosine and uracil, are shown at the bottom. (http://tonga.usip.edu/gmoyna/biochem341/lecture36.html)

Due to its relative instability, single-stranded DNA rarely exists as shown in figure 1.2c. In fact the only time that single-stranded DNA is exposed in living cells is during active processing, i.e. when replication, transcription or repair are occurring. At all other times DNA exists as a double helix containing two antiparallel strands of single-stranded DNA. As previously stated, double-stranded DNA is capable of adopting several different forms depending on the salt and water content of the sample. In the biologically-relevant B form of double-stranded DNA each turn is approximately 34 Å long and contains 10 bases (Watson and Crick, 1953). The stability of hydrogen bonding between the four canonical bases of DNA largely defines the interactions that occur between antiparallel strands, such that guanine invariably base pairs with cytosine through the formation of 3 hydrogen bonds, while

adenine pairs with thymine through the formation of 2 such bonds. This process is known as Watson-Crick base pairing and is demonstrated below in figure 1.4b.

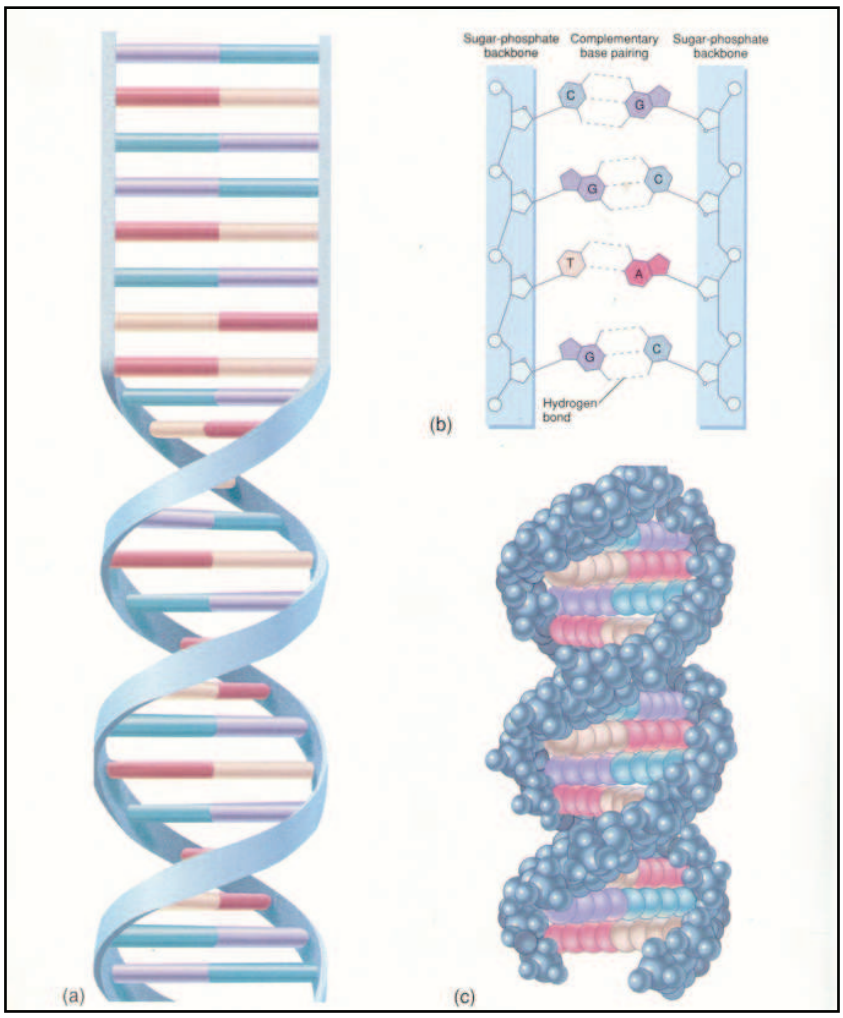


Figure 1.4: Three representations of the double helical structure of DNA and the Watson-Crick pairing that defines the interactions between opposing bases. In all diagrams adenine is indicated in pink, cytosine in blue, guanine in purple and thymine in beige. The hydrophilic sugar-phosphate backbone is also shown in blue and can be seen in all 3 diagrams wrapping around the internalised, hydrophobic bases (Saladin, 1998).

Due to the asymmetry of nucleotides, base pairs are unable to occupy the same plain. As such, in the standard B-conformation of double-stranded DNA there is always slightly more space on one side of base-paired nucleotides than the other. The net consequence of this is that B-form DNA possesses a minor and a major groove, as

shown in figure 1.5, with the major groove occupying roughly 50 % more space than the minor. Both grooves are of biological significance as they provide the binding sites for numerous enzymes and chemical compounds, including a number of notable carcinogens (Cosman *et al.*, 1992).

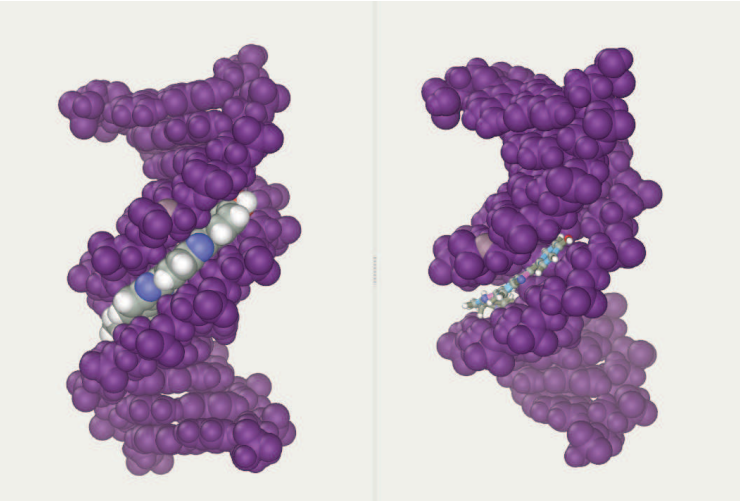


Figure 1.5: Double-stranded DNA containing the double-stranded DNA-specific dye Hoescht 33258 bound to the minor groove.
(http://commons.wikimedia.org/wiki/File:DNA-ligand-by-Abalone.png)

1.3: Function of DNA

As previously stated, DNA provides the mechanism by which the vast majority of life on earth is able to propagate and maintain its biological make-up over successive generations. It provides the basic unit of heredity for all living organisms with the exception of certain viruses. It is through the high fidelity replication of chromosomal DNA, which occurs every time a cell divides, that organisms are capable of maintaining their genetic composition over time and of conferring this composition to their offspring.

The process of DNA replication is often referred to as semi-conservative since each of the two strands of parental genomic DNA functions as a template for replication of a

daughter strand. The two resulting genomes therefore contain one parental strand and one daughter strand each. As was originally proposed by Watson and Crick (Watson and Crick, 1953), the base pairing that defines the interactions between two strands of DNA also defines the sequence of nucleotides that are incorporated into growing DNA strands during DNA replication. Watson and Crick's model of DNA replication relied on the fact that adenine is invariably incorporated opposite template strand thymine and thymine opposite adenine. Similarly, guanine is invariably incorporated opposite cytosine and vice versa. This model of replication was later confirmed to be true by Meselson and Stahl (Meselson and Stahl, 1958).

As part of its role in heredity, DNA also encodes virtually all of the proteins that are required for normal biological function. These proteins are expressed through the combined actions of exceptionally complicated and highly regulated mechanisms in two processes, known as transcription and translation. Transcription of DNA is initiated by binding of transcription factors to the promoter region of the gene. The promoter typically contains a number of conserved regulatory sequences, such as the eukaryotic/archaeal TATA box or the bacterial Pribnow box (Wobbe and Struhl, 1990; Pribnow, 1975). Once the necessary regulatory criteria have been fulfilled, an RNA polymerase then binds to the promoter region and proceeds to transcribe the gene. Transcription results in the template-dependent production of messenger RNA (mRNA), which is complementary to the DNA sequence of the gene except uracil replaces thymine opposite adenine. Once fully transcribed, the messenger RNA sometimes requires processing in eukaryotes to remove non-coding regions called introns through a process known as splicing. The mature mRNA is then translated by a ribosome, which uses the mRNA as a template to catalyse the polymerisation of amino acids, via transfer RNA molecules (tRNAs), into a polypeptide. In eukaryotes further post-translational modifications, such as glycosylation and phosphorylation, of the polypeptide chain are commonly made before the protein is ready to perform its biological role.

1.4: The three domains of life

The variation that exists between and within different species is primarily due to differences in the sequence of bases that make up the genomes of living organisms. During the twentieth century DNA and RNA sequencing technologies emerged, enabling evolutionary biologists to investigate the genetic make-up of different organisms for the first time and gain a unique insight into the true phylogeny of life on earth. Up until that point the putative classification of living organisms involved four, five or even six kingdoms separated on the basis of classical phenotypes (Whittaker, 1969). During the 1970s, however, Carl Woese and colleagues performed a number of pioneering studies focusing on sequence homology in 16S ribosomal RNA (rRNA). These studies led to the establishment of a new taxon, above the level of kingdom, and hence the division of life on earth into three domains: the Archaea, the Bacteria and the Eukarya (Woese *et al.*, 1990).

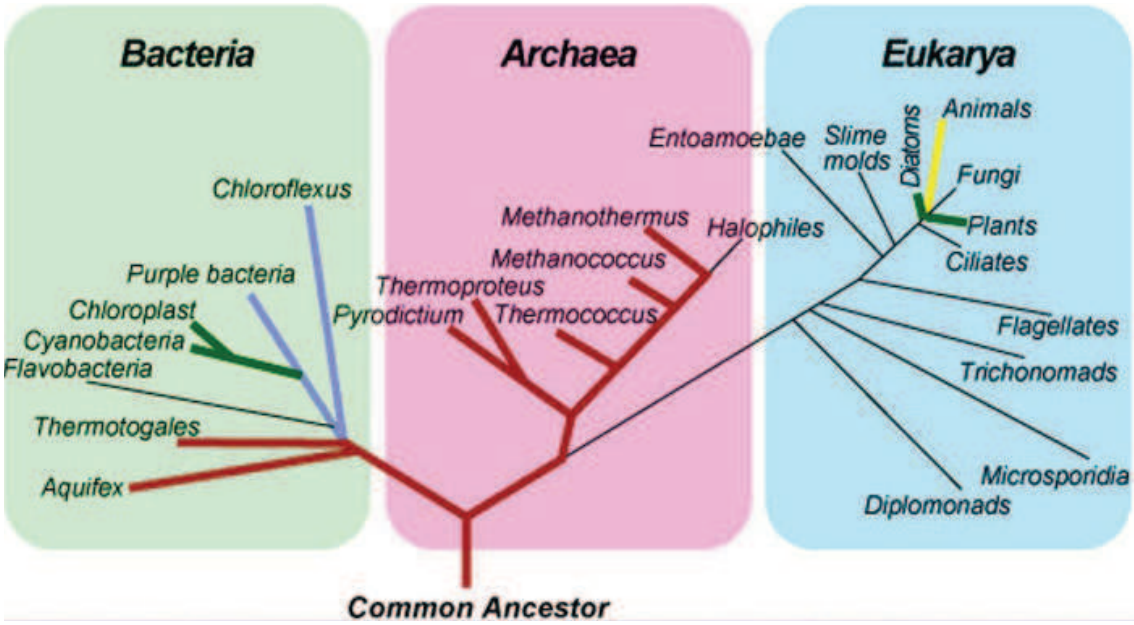


Figure 1.6: A phylogenetic tree showing the three domains of life all originating from a single common ancestor.
(http://www.learner.org/courses/envsci/visual/visual.php?shortname=common_ancestor)

1.5: Archaea as a model for eukaryotic systems

To cellular biologists the division between single-celled prokaryotic organisms and eukaryotes was always explicit. However, the advent of molecular biology led to the taxonomic subdivision of prokaryotes into archaea and bacteria based on profound molecular differences. The principal discoverer of the archaeal domain, Carl Woese, himself noted "On the cytological level archaebacteria are indeed prokaryotes (they show none of the defining eukaryotic characteristics), but on the molecular level they resemble other prokaryotes, the eubacteria, no more (probably less) than they do the eukaryotes" (Woese *et al.*, 1990).

Since Woese's seminal findings, it has become increasingly apparent that much of the cellular machinery of archaea is highly homologous to that of eukaryotes; albeit usually more rudimentary in its nature (Tye, 2000). Due to the greater complexity of eukaryotic systems and the inherent difficulties of working with mesophilic proteins, thermophilic archaea have become ideal candidates for use in *in vitro* biochemical analyses owing to their intrinsic structural stability (Bell *et al.*, 2005).

The archaeal domain currently consists of several phyla, including the Korarchaeota, Nanoarchaeota and Thaumarchaeota, however, the Crenarchaeota and Euryarchaeota represent by far the best-characterised of all the archaeal phyla. The *Sulfolobus* genus of the crenarchaea, which includes *S. solfataricus* and *S. islandicus*, is perhaps the best characterised of all the archaeal phyla and has been the subject of numerous studies (Brock *et al.*, 1972; She *et al.*, 2002; Robinson *et al.*, 2004). Unusually among extremophiles, *S. solfataricus* is aerobic and grows chemoorganotropically, making it an ideal model organism for culturing in a laboratory setting. The euryarchaeal phylum also represents a major focus of interest among researchers (Greagg *et al.*, 1999; Ishino and Ishino, 2001; Henneke *et al.*, 2005) and includes the well characterised *Pyrococcus* and *Thermococcus* genera. The euryarchaeal phylum is of particular interest to clinical researchers as it includes methanogens, which inhabit the gastrointestinal tracts of humans (Rajilić-Stojanović *et al.*, 2007). The main focus of this thesis will be the mechanisms of DNA damage recognition and response by DNA polymerases in the Euryarchaeota.

1.6: DNA polymerases

1.6.1: Overview

Shortly after Watson and Crick's seminal paper, a flurry of related research took place leading to the discovery of the first DNA polymerase, DNA polymerase I (Pol I) from *Escherichia coli*, in 1956 (Kornberg *et al.*, 1956). Arthur Kornberg and colleagues demonstrated that DNA polymerases function by catalysing the addition of deoxyribonucleoside triphosphates (dNTPs) to the 3' OH- group of growing DNA strands. The process of deoxynucleotide incorporation is usually carried out by a polymerase while reading from a template polynucleotide strand in a process known as semi-conservative replication (Meselson and Stahl, 1958; Lehman *et al.*, 1958), as originally proposed by Watson and Crick (Watson and Crick, 1953).

1.6.2: DNA polymerase structure

While sequence homology across all DNA polymerases is relatively low, their overall structure shows a high degree of conservation across all domains (Rothwell and Waksman, 2005). The first crystal structure to be solved of a DNA polymerase was that of the large proteolytic Klenow fragment of *E. coli* Pol I (Ollis *et al.*, 1985). The Klenow fragment contains the polymerase and 3'-5' exonuclease domains but lacks the C-terminus, in which the 5'-3' exonuclease domain resides. The crystal structure revealed three core features known as the fingers, palm and thumb domains, as shown in figure 1.7. Unsurprisingly, the names of the three domains refer to the overall structure of polymerases, which tends to resemble that of an open right hand. The functional role of each domain appears to be consistent across almost all known polymerases. The thumb domain is known to be involved in binding double-stranded DNA, as well as translocation and processivity. The palm domain is responsible for catalysing the phosphoryl transfer reaction while the fingers are responsible for

selectively binding incoming dNTPs (Rothwell and Waksman, 2005). The fingers domain has also been shown to interact with the template base, to which the incoming dNTP pairs (Steitz, 1999).

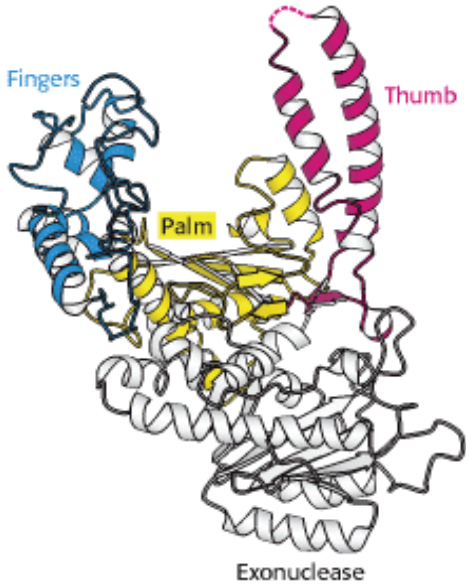


Figure 1.7: The tertiary structure of the Klenow fragment of *E. coli* Pol I, as solved by X-ray crystallography (Berg *et al.*, 2006).

As previously stated, almost all known polymerase structures show striking similarity in three core features; the palm, fingers and thumb domains. Reference to figure 1.8 reveals the conservation of these features across several of the main families of polymerases, as revealed by X-ray crystallographic studies. However, whether or not these core features are truly homologous is open to debate (Steitz *et al.*, 1994).

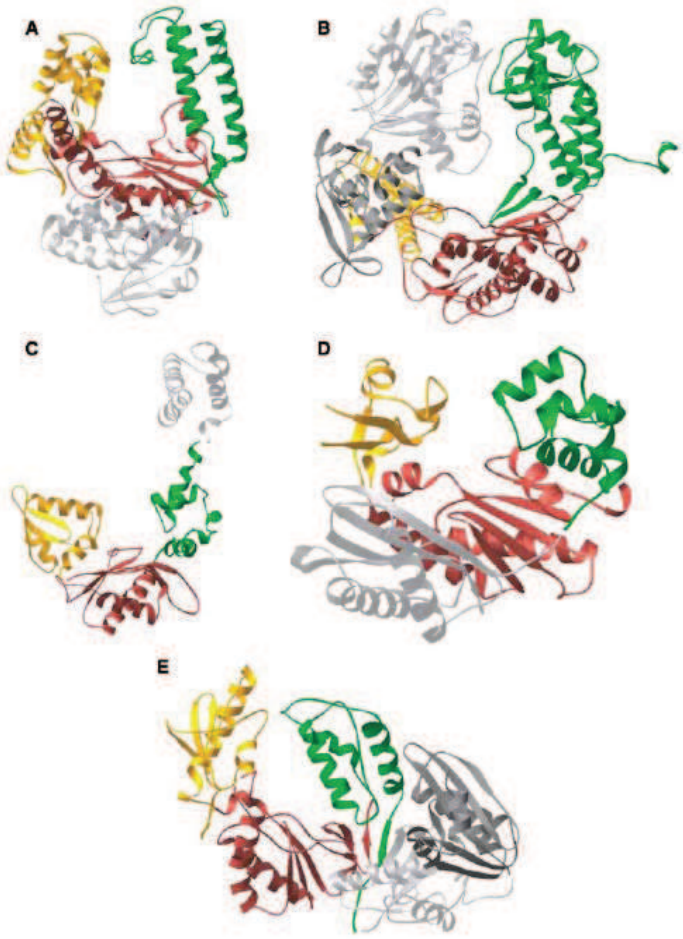


Figure 1.8: Tertiary structures of family A, B, X, Y and RT DNA polymerases, as revealed by X-ray crystallography. The fingers, palm and thumb domains are indicated in yellow, red and green, respectively. **A)** Apo Klentaq1 (family A). **B)** Apo RB69 DNA polymerase (family B). **C)** Apo Pol β DNA polymerase (family X). **D)** Dpo4 DNA polymerase (family Y). **E)** The p66 subunit of reverse transcriptase (RT family) (Rothwell and Waksman, 2005).

1.6.3: DNA polymerase function

1.6.3.1: Overview

The majority of DNA polymerases function by catalysing the addition of deoxyribonucleotides to the 3' ends of growing DNA strands in a template-dependent

manner. The functional conservation of this process is fairly evident across most polymerases and can be separated into five main steps, as shown diagrammatically in figure 1.9 (Rothwell and Waksman, 2005).

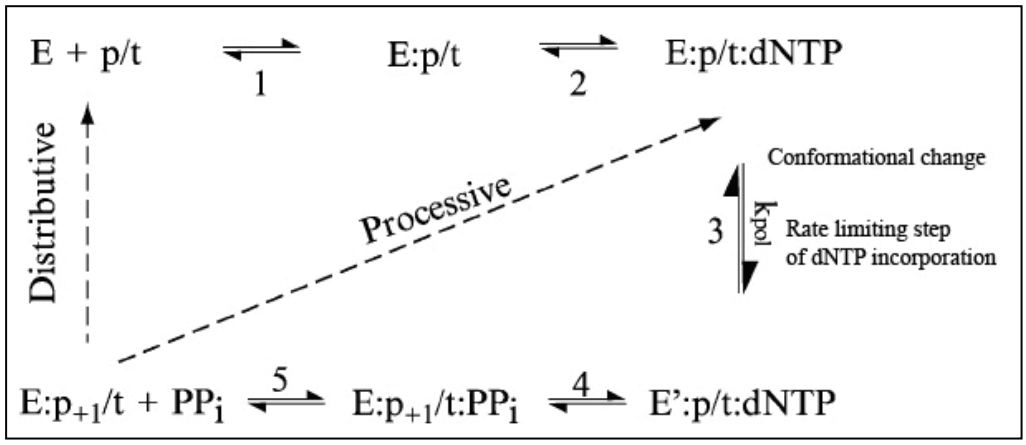


Figure 1.9: A summary of the 5 main steps in the kinetic pathway of nucleotide incorporation. The names and designations of each of the enzyme/DNA/dNTP complexes are referred to in the text. The rate constant of the conformational change, k_{pol}, which forms the rate-limiting step in nucleotide incorporation, is also indicated at step 3. (Rothwell and Waksman, 2005).

The first step in the dNTP incorporation pathway involves binding of a DNA primer-template (p/t) by the polymerase (E) to form an enzyme-DNA complex (E:p/t); the thumb domain is thought to play a crucial role in this step. Next, the fingers domain of the polymerase selectively binds an incoming dNTP via a conserved group of positively charged amino acids to form an enzyme-DNA-dNTP ternary complex (E:p/t:dNTP). Selective binding of the dNTP is dependent on divalent cations to stabilise the complex and results in a significant conformational change by the polymerase whereby the fingers domain undergoes a large rotation to form a closed catalytically active complex (E':p/t:dNTP). This conformational change is believed to be the rate-limiting step in nucleotide incorporation. Fourthly, nucleophilic attack by the 3'-OH primer terminus on the 5' α phosphate of the incoming dNTP follows. The attack occurs within the catalytic palm domain, resulting in the formation of a phosphodiester bond (as shown in figure 1.10). The step also results in production of a pyrophosphate product ($E:p_{+1}/t:PP_i$). Finally, a second conformational change takes

place resulting in release of the pyrophosphate product ($E:p_{+1}/t + PP_i$). At this point either the polymerase dissociates from the primer-template, in the case of distributive synthesis, or the primer terminus is translocated in preparation for another round of nucleotide incorporation, as occurs in processive synthesis (Rothwell and Waksman, 2005).

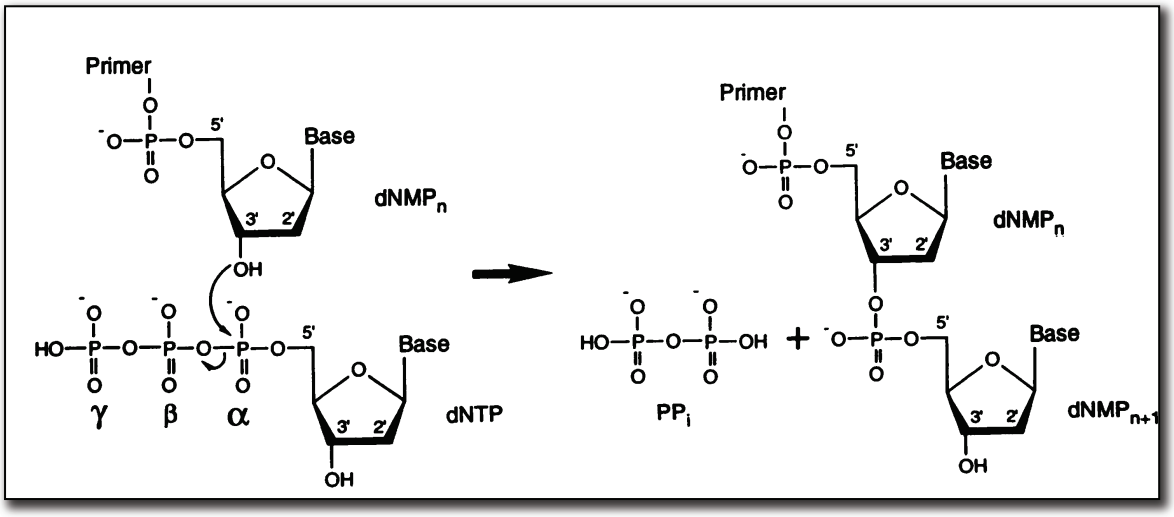


Figure 1.10: Nucleophilic attack by the 3' OH group of a DNA primer (shown at the top of the diagram) on the α phosphate of an incoming dNTP (shown at the bottom), resulting in the formation of a phosphodiester bond between the two nucleotides and release of pyrophosphate (Pelletier *et al.*, 1994).

1.6.3.2: Two metal ion requirement in polymerisation

Several of the key steps in the nucleotide incorporation pathway were first proposed by Burgers and Eckstein based on studies using phosphorothioate dATP analogs (Burgers and Eckstein, 1979). The pair showed that during polymerisation a divalent metal ion binds to the β and γ phosphates of the dNTP in a β,γ-bidentate manner. They also showed that the negative charge on the α phosphate of the incoming dNTP is neutralised by a positive group on the polymerase. Nucleophilic attack by the primer's 3' -OH group on the α phosphate and release of PP_i was also shown to proceed in an in-line fashion with inversion of configuration at phosphorus.

Based on X-ray crystallographic studies, the process of nucleotide incorporation is now known to be dependent upon binding of two divalent metal ions, usually Mg^{2+}, to promote the stability of the polymerase:primer/template:dNTP complex in the transition state (Pelletier *et al.*, 1994). This two-metal-ion model was first suggested based on studies of the better characterised 3'-5' exonucleolytic pathway of the Klenow fragment of *E. coli* Pol I (as described in chapter 1.5.3.3) and was first demonstrated using rat DNA polymerase beta (β). The rat Pol β enzyme was captured in a transition state ternary complex with a DNA primer-template and a ddCTP molecule. The interaction was shown to involve three key carboxylate residues, asp190, asp192 and asp256, as well as two hexacoordinated, partially hydrated metal ions, which are separated by about 4 Å (Pelletier *et al.*, 1994; Sawaya *et al.*, 1997). Subsequent sequence analyses and structural studies revealed that in fact only two of the carboxylate residues, asp190 and asp256, are functionally conserved. The interactions of the aspartic acid residues and metal ions with the DNA primer and an incoming ddCTP molecule are shown in figure 1.11.

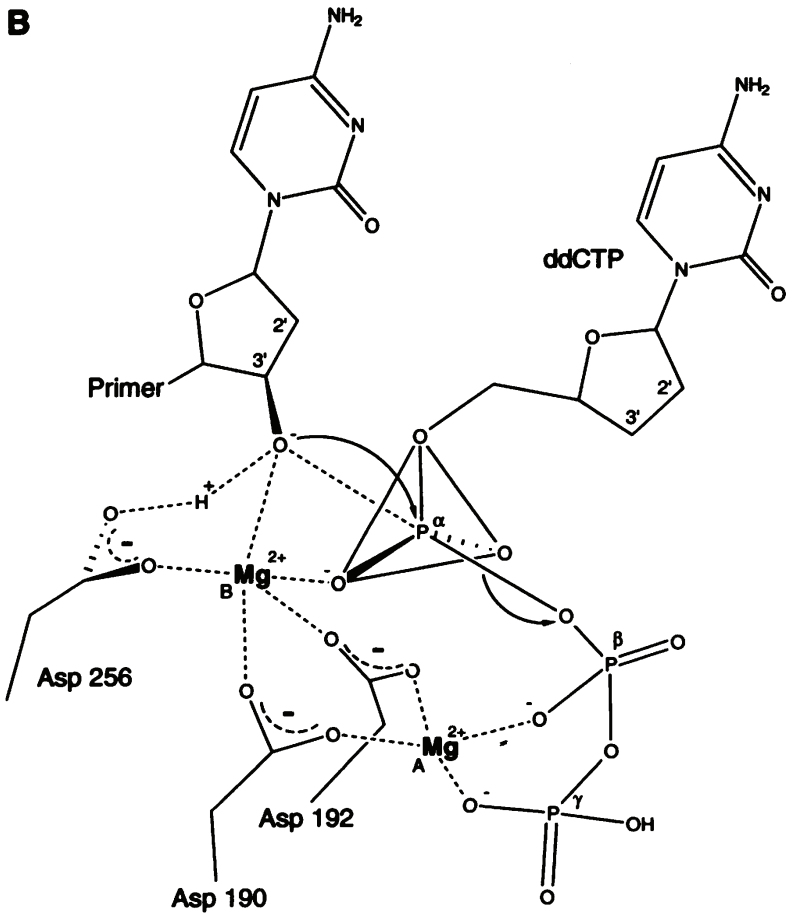


Figure 1.11: Interactions of three key carboxylate residues of rat Pol β with the primer DNA strand and incoming ddCTP during nucleotide incorporation. The diagram shows the critical roles played by two hexacoordinated, partially-hydrated Mg^{2+} ions in stabilising the transition state complex (Pelletier *et al.*, 1994).

Further structural studies revealed that the two metal ion mechanism of nucleotide incorporation appears to be universal across polymerases. For example, the equivalent residues, D705 and D882 in *E. coli Pol* I, play a similar role in Mg^{2+}-binding to the asp190 and asp256 residues of rat Pol β, as shown in figure 1.12 (Steitz, 1999). With reference to this image, it is thought that the interaction of metal ion B with the 3' -OH group of the primer strand serves to lower the acid dissociation constant (pK_a) of the hydroxyl group thereby facilitating its attack on the α-phosphate of the incoming dNTP. Metal ion A binds to, and is assumed to facilitate the separation of, the β and γ phosphates of the dNTP. It has been proposed that both metal ions play a role in stabilising the pentacovalent transition state of the ternary

complex. Mutagenic analysis, for example of the conserved YGDTDS region of eukaryotic Pol α, has also been used to demonstrate the functional role of aspartic acid-containing motifs in nucleotide binding and incorporation (Copeland and Wang, 1993).

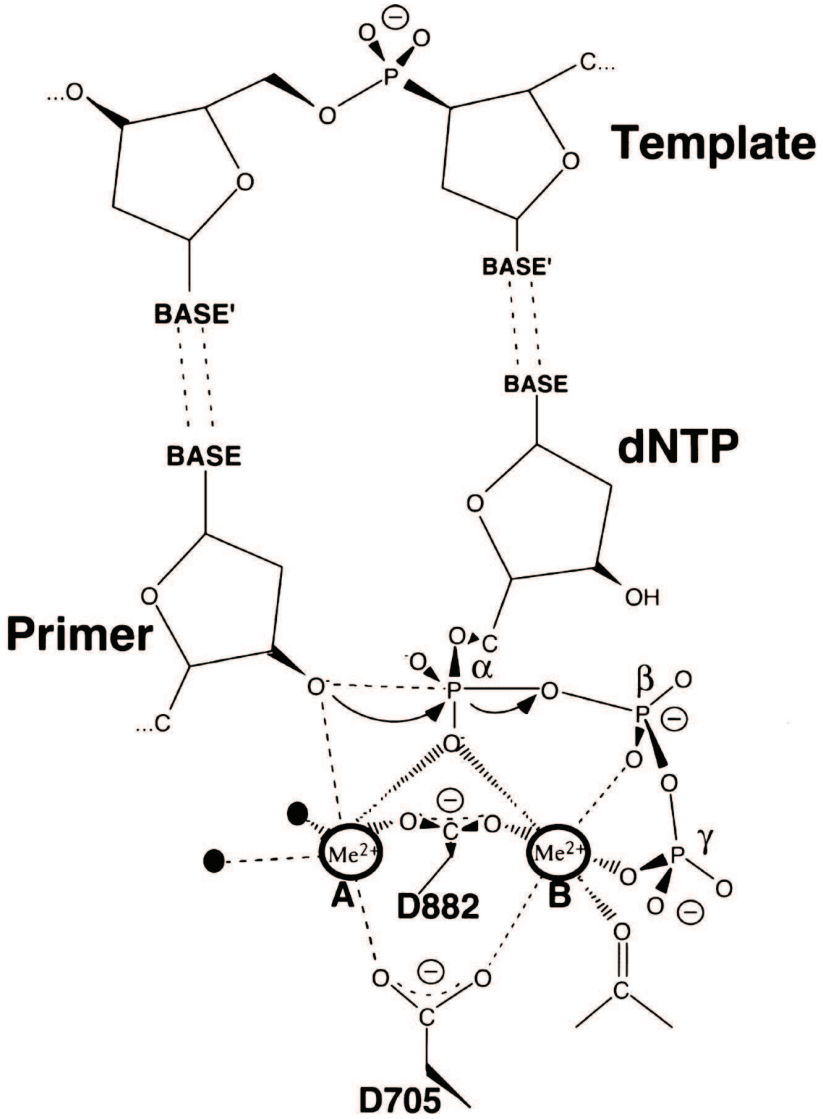


Figure 1.12: Interactions of two conserved aspartic acid residues, D705 and D882, in the Klenow fragment of *E. coli* Pol I with the DNA primer strand and incoming dNTP during nucleotide incorporation. The roles of two metal ions, A and B, in stabilising the transition state ternary complex are also shown (Steitz, 1999).

1.6.3.3: Fidelity and 3'-5' exonucleolytic activity

3'-5' exonucleolysis is a hydrolytic reaction, which results in cleavage of the phosphodiester bond nearest the exposed 3' end of DNA strands and causes release of a deoxyribonucleoside 5'-monophosphate (Kunkel, 1988). 3'-5' exonucleolytic activity is also often referred to as "proofreading" exonucleolytic activity as it serves to increase fidelity (defined as the number of errors per base incorporated) during replication by catalysing the cleavage of terminal nucleotides from polynucleotide chains; usually in response to base misincorporation. Whilst the majority of polymerases exhibit extremely high levels of fidelity, with error rates ranging from 10^{-4} to 10^{-7} (Kunkel and Bebenek, 2000), generally speaking, replicative polymerases that possess 3'-5' exonuclease activity exhibit the highest fidelities.

As previously stated, just like nucleotide incorporation, the process of 3'-5' exonucleolysis in DNA polymerases also relies upon binding of two divalent metal ions. This two metal ion system was first demonstrated using the Klenow fragment of *E. coli* Pol I (Beese and Steitz, 1991). Beese and Steitz showed that the Klenow fragment adopts a transition state ternary complex containing the enzyme, a single-stranded DNA substrate and a dTMP product. The transition state complex was shown to require stabilisation by interactions with two metal ions, as shown in figure 1.13. The two metal ions, referred to as A and B, are usually Mg^{2+} and are separated by a distance of 3.9 Å in the Klenow fragment of Pol I. Mn^{2+} and Zn^{2+} have also been shown to be interchangeable with Mg^{2+} in stabilising the transition state complex (Beese and Steitz, 1991).

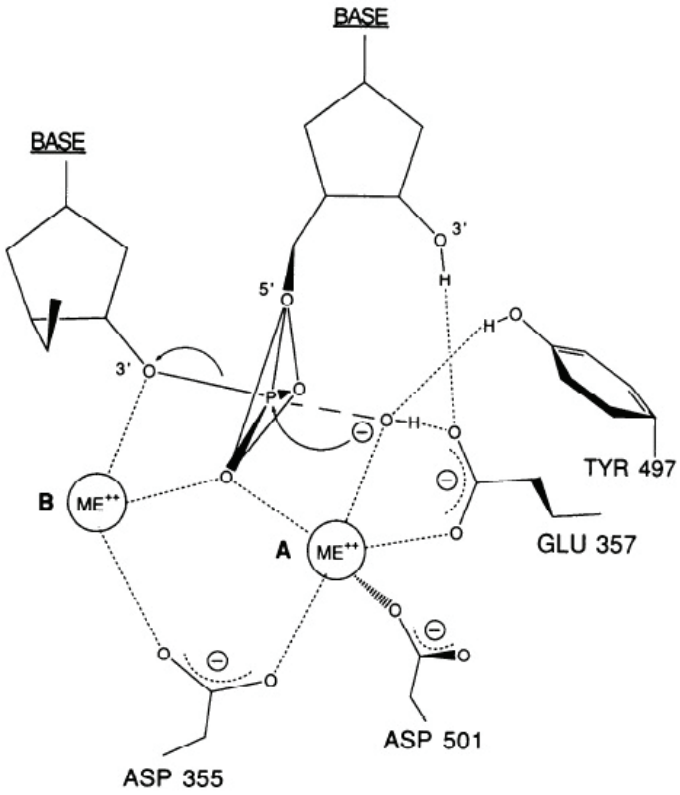


Figure 1.13: The transition state of the Klenow fragment of *E. coli* Pol I during 3'-5' exonucleolysis. The interactions of 4 key residues with two divalent metal ions and a single-stranded DNA substrate are indicated (Beese and Steitz, 1991).

It has previously been demonstrated that replicative DNA polymerases appear to adopt two distinct conformations in order to facilitate switching between polymerisation and 3'-5' exonucleolytic activity (Freemont *et al*, 1988; Baker and Bell, 1998). Based on the results of X-ray crystallographic studies, Freemont and colleagues observed that when bound to a DNA primer-template the large proteolytic Klenow fragment of *E. coli* Pol I appears to adopt either a polymerising conformation, as occurs under normal circumstances, or an editing conformation, which results from base misincorporation. In the editing conformation the polymerase appears to shift the 3' terminus of the primer towards the exonuclease domain, thereby promoting removal of the misincorporated base, as shown in figure 1.14. By stochastically favouring degradation of the primer, the polymerase is subsequently afforded another opportunity to insert the correct nucleotide. Since this original finding, various studies have shown that adoption of an editing conformation following base

misincorporation is a fairly universal response of replicative DNA polymerases (Joyce and Steitz, 1994; Datta *et al.*, 2009). Replicative polymerases also extend mismatches very inefficiently thus increasing the likelihood of the enzyme adopting an editing conformation and exhibiting proofreading exonucleolytic activity.

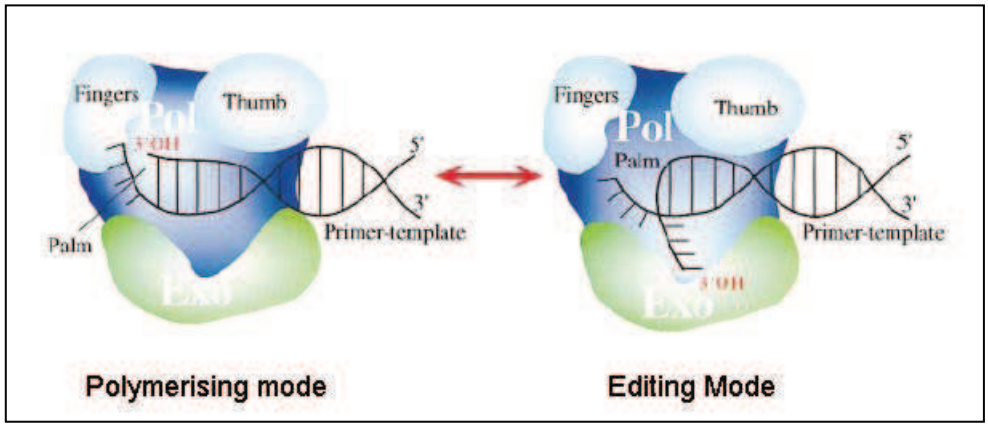


Figure 1.14: Representations of the polymerising and editing modes of the Klenow fragment of *E. coli* Pol I (Baker and Bell, 1998)

Interestingly, it has been suggested that shuttling of the 3' primer terminus between the polymerase and exonuclease domains not only serves to promote proofreading exonuclease activity, but also occurs in response to binding of correct or incorrect dNTPs (Datta *et al.*, 2010). In doing so, the exonuclease domain functions as a pre-catalytic checking mechanism that helps to prevent base misincorporation and increase fidelity.

1.6.3.4: Primases

DNA polymerases are, in themselves, incapable of synthesising DNA *de novo*; instead they require a short RNA primer of approximately 10 bases from which they can extend. A specialised primase therefore performs the role of primer synthesis in all three domains of life. The *E. coli* primase exists as a monomer known as DnaG, while in eukaryotes the p48 and p58 subunits of the heterotetrameric DNA polymerase alpha (α)–primase complex perform this role. Initially, it was thought that archaea differ significantly from the other two domains in that its p41 primase,

which appears to be homologous to the eukaryotic p48 subunit, does not require an RNA primer. Instead, p41 preferentially incorporates deoxynucleotides and can synthesise several kilobase long fragments of DNA *de novo* (Bocquier *et al.*, 2001). Since then, p41 has been shown to associate with a 46 kDa subunit (p46), which is homologous to the eukaryotic p58 primase subunit. The combined activity of p41:p46 appears to suggest that, as in the other two domains, archaeal primase also initiates replication through the production of an RNA primer (Liu *et al.*, 2001; Matsui *et al.*, 2003).

In eukaryotes, after synthesis of the RNA primer, DNA polymerase α performs the role of primer elongation. It incorporates several hundred bases before the replicative polymerases, Pol δ and ε, take over and perform the majority of chromosomal DNA replication (Garg and Burgers, 2005; Nick McElhinny *et al.*, 2008; Pursell *et al.*, 2007). Interestingly, while the archaeal p41-p46 primase possesses reduced processivity compared to the p41 subunit alone, it is still capable of synthesising several hundred base fragments of DNA without dissociating, indicating that the enzyme plays a much greater role in DNA replication than its eukaryotic counterpart (Liu *et al.*, 2001). Given the lack of a DNA polymerase α homologue in archaea, it seems likely that the dimeric p41:p46 primase enzyme performs the roles of both primer synthesis and elongation in this domain (Barry and Bell, 2006). Following the production and elongation of an RNA primer, archaeal chromosomal DNA replication, as in the other two domains, is thought to proceed through the combined actions of replicative polymerases (Henneke *et al.*, 2005). The process of chromosomal DNA replication involves numerous other partners, a number of which will be described in chapter 1.6.5.

1.6.3.5: Replicative and repair polymerases

The majority of known DNA polymerases are described as replicative and, unsurprisingly, are primarily responsible for replication of chromosomal DNA. While replicative polymerases perform an extremely efficient job of amplifying DNA with high fidelity and processivity (defined as the average number of nucleotides incorporated per binding/dissociation event), they tend to stall replication upon encountering DNA lesions, such as abasic (AP) sites. Compared to prokaryotes, there

exists a much larger number of specialised polymerases in the Eukarya. These additional enzymes tend to enable efficient processing or bypass of specific DNA lesions and are therefore known as repair or translesion synthesis (TLS) polymerases (Plosky and Woodgate, 2004). Notable examples of eukaryotic repair polymerases include Pol β, which is known to function primarily in the base excision repair pathway (Idris *et al.*, 2001) and Pol ζ, which is responsible for synthetic bypass of a variety of lesions (Gan *et al.*, 2008). These repair polymerases appear to have evolved to counter situations where low-fidelity bypass of a lesion is preferable to strand termination and cell death (Rattray and Strathern, 2003). To enable them to deal with abnormal adducts and lesions, repair polymerases generally exhibit much lower fidelity than their replicative counterparts (Plosky and Woodgate, 2004). Unsurprisingly, they also exhibit significantly reduced processivity compared to replicative polymerases, since additional low-fidelity replication beyond the lesion would be undesirable from an evolutionary perspective (Friedberg *et al.*, 2002).

As well as replicative and repair DNA polymerases, other more specialised examples exist in higher eukaryotes; for example, DNA polymerase mu (μ), which is part of the X-family of polymerases and has been implicated in the somatic hyper-mutation of immunoglobulin genes (Dominguez *et al.*, 2000). The family X polymerases in general display unusual behaviours and have been shown to be capable of producing atypical DNA structures. It has been suggested that these unusual structures might function as checkpoints or "flag-wavers" for members of different DNA repair pathways (Ramadan *et al.*, 2004). Telomerases represent another highly specialised form of polymerase and are capable of extending the telomeric ends of eukaryotic chromosomes in a DNA template-independent manner. Rather than utilise a deoxynucleotide template, as is the case with most DNA polymerases, telomerases instead employ an internalised RNA sequence located within the enzyme (Mason *et al.*, 2010).

1.6.3.6: Additional activities of DNA polymerases

In addition to catalysing the incorporation of deoxynucleotides into polynucleotide chains, DNA polymerases possess a variety of other activities, some of which have already been discussed. Bacterial Pol I, for example, possesses a 5'-3' exonuclease

activity (also referred to as a flap endonuclease activity), which enables it to process Okazaki fragments into mature DNA strands (as described in chapter 1.6.5). Many polymerases also possess a 3'-5' (proofreading) exonuclease domain enabling them to cleave nucleotides from the 3' ends of DNA strands, as described in chapter 1.6.3.4.

Repair enzymes, like DNA Pol ß and theta (θ), possess deoxyribosephosphate (dRP) lyase activity, which facilitates the excision of damaged bases from DNA strands by cleavage of the 5' sugar phosphate group in polynucleotide chains (Allinson *et al.*, 2001; Prasad *et al.*, 2009). The dRP lyase activities of these enzymes therefore play an essential role in the base excision repair pathway. Another activity can be found in reverse transcriptases, for example that of the human immunodeficiency virus (HIV), and is referred to as RNAseH activity. It enables the specific degradation of RNA strands in double-stranded DNA/RNA hybrids, using a two metal ion mechanism, to promote the production of double-stranded DNA (Freed and Martin, 2001).

1.6.4: Classification of DNA polymerases

During the early stages of polymerase research the majority of studies focused on enzymes derived from bacteria, such as *E. coli* and *Thermus aquaticus* (Ollis *et al.*, 1985; Chien *et al.*, 1976). Whilst universal traits have been found in all DNA polymerases, such as the overall structure described in chapter 1.6.2, significant differences have also been discovered within and between the three domains. With the expansion of research from bacterial polymerases into other model organisms, such as the bacteriophage RB69 (Shamoo and Steitz, 1999), archaea like *Pyrococcus furiosus* (Lundberg *et al.*, 1991) and more recently eukaryotic systems like the baker's yeast *Saccharomyces cerevisiae* (Burgers, 1998), a logical system of classification has now emerged.

On the basis of sequence homology (Delarue *et al.*, 1990; Ito and Braithwaite, 1991; Braithwaite and Ito, 1993) and crystal structural data (Joyce and Steitz, 1994), the most recent classification of DNA polymerases comprises seven families: A, B, C, D, X, Y and RT. A summary of the main features of each family is provided in table 1.1.

The main focus of this thesis will be the two putative replicative polymerases of the euryarchaea; the family B and D polymerases.

Family	Domains	Examples	Primary role
A	Viruses	T7	Replicative
	Bacteria	Pol I, *Taq*	Okazaki fragment processing/ excision repair
	Eukarya	Pol γ	Replicative (mitochondrial DNA)
B	Viruses	T4, RB69	Replicative
	Archaea	Pfu-Pol B, Pab-Pol B	Replicative?
	Eukarya	Pol α , Pol δ, Pol ε, Pol ζ	Replicative and repair
C	Bacteria	Pol III	Replicative
D	Euryarchaeota	Pfu-Pol D	Replicative?
	Nanoarchaeota	Neq-Pol D	Replicative?
	Korarachaeota	Kcr-Pol D	Replicative?
	Thaumarchaea	Nma-Pol D	Replicative?
X	Eukarya	Pol ß	Repair (base excision)
		Pol λ, Pol μ	Repair (double-stranded breaks)
		Terminal deoxynucleotidyl transferase (TdT)	Generation of immunological diversity
Y	Bacteria	Pol IV, Pol V	Repair (translesion synthesis)
	Archaea	DinB homologues	Repair (translesion synthesis)
	Eukarya	Pol ι, Pol κ	Repair (translesion synthesis)
RT	Retroviruses	HIV reverse transcriptase	Proliferation/infection
	Eukarya	Telomerase	Telomere maintenance

Table 1.1: Summaries of the seven families of DNA polymerases. The domains in which each family is found are listed in column 2. Examples from each domain and the primary role that each of them plays are also provided.

1.6.4.1: Family B polymerases

The B family of DNA polymerases are present in all archaea and eukaryotes and includes a number of viral examples. Almost all family B polymerases are defined as replicative and possess 3'-5' exonuclease activity. Indicative of their principal role in chromosomal DNA replication is the high fidelity that they tend to exhibit. This fidelity is facilitated by a remarkably strong 3'-5' exonuclease activity; over 1000 times higher than that of *E. coli* Pol I (Capson *et al.*, 1992; Lin *et al.*, 1994). Notable examples of viral family B polymerases include the well-characterised bacteriophage T4 and RB69 polymerases. Well-known archaeal family B polymerases include those of *Pyrococcus furiosus* (Pfu), *Thermococcus gorgonarius* (Tgo) and *Thermococcus kodakarensis* (Tko). Four family B DNA polymerases are found in eukaryotes, including the replicative polymerases α, delta (δ) and epsilon (ε), as well as the repair polymerase, Pol zeta (ζ).

Whilst there is still considerable doubt about the roles of archaeal DNA polymerases, it is thought that the archaea generally possess two functional, replicative polymerases. In the case of crenarchaeal organisms one or two family B polymerases may perform the role of chromosomal DNA, however, a third family B polymerase is also found in numerous species (Edgell *et al.*, 1997). In euryarchaeal organisms, on the other hand, it is thought that chromosomal DNA replication is performed by the combined actions of one family B and one family D polymerase (Henneke *et al.*, 2005). Interestingly, very little is known about why three family B polymerases are found in certain members of the crenarchaea or what role each of them plays (Duggin and Bell, 2006). Based on sequence analyses, at least one of these 3 family B enzymes appears to be functionally inactive with regard to polymerisation activity (Rogozin *et al.*, 2008). Interestingly, this catalytically inactive polymerase is also found in members of the euryarchaea. It has been speculated, due to the high degree of conservation shown throughout the archaea, that these inactive enzymes may play some sort of structural role, functioning as supportive elements in the overall replication machinery (Rogozin *et al.*, 2008; Tahirov *et al.*, 2009). Certainly the more complex, multimeric eukaryotic polymerases appear to have evolved in this way.

Despite suggestive biochemical and bioinformatic analyses, it should be stressed that the assignment of the family B DNA polymerases in archaea as replicative awaits rigorous determination. Much stronger evidence, based on genetic studies, shows that Pol III is replicative in bacteria (Kelman and O'Donnell, 1995; O'Donnell, 2006) and Pols δ and ε in eukaryotes (Garg and Burgers, 2005; Nick McElhinny *et al.*, 2008; Pursell *et al.*, 2007).

1.6.4.2: Family D polymerases

D family polymerases are perhaps the least well characterised of all the polymerases but are thought to play a replicative role (Uemori *et al.*, 1997). All known examples of this family are found in the euryarchaeal, nanoarchaeal, korarchaeal and thaumarchaeal phyla of the archaea. The family D polymerases are heterodimeric enzymes comprising a large subunit of approximately 140 kDa, which is responsible for polymerase activity, and a smaller subunit of approximately 70 kDa, on which the 3'-5' exonuclease domain resides (Uemori *et al.*, 1997). There is some evidence to suggest that the heterodimeric enzyme itself forms homodimers, but whether or not this association is functional is open to debate (Shen *et al.*, 2001). Interestingly, the smaller subunit shows small, but significant, homology to the p50 subunit of eukaryotic Pol δ, providing further evidence for the conservation of cellular mechanisms between the two domains (Cann *et al.*, 1998).

As previously stated, it is believed that the functional euryarchaeal family B and D polymerases play replicative roles. Based on their respective strand displacement activities, it has been suggested that there is a division of labour between the family B and D polymerases of euryarchaea with the family B polymerase proposed to be responsible for leading strand replication and the family D polymerase responsible for lagging strand replication in a manner analogous to the leading/lagging strand model of eukaryotic DNA replication (Henneke *et al.*, 2005), as described in chapter 1.6.5. It should be noted, however, that biochemical evidence can never conclusively assign replication of the leading and lagging strands to a particular polymerase.

1.6.4.3: Family Y polymerases

Y family polymerases are found in eubacteria, archaea and eukaryotes and are primarily involved in translesion synthesis; particularly bypass of lesions associated with UV damage (Goodman, 2002). As repair enzymes, they exhibit low fidelity; in the range of 10^{-3} - 3 x 10^{-5} errors/base replicated (Johnson *et al.*, 2000; Ohashi *et al.*, 2000, Tang *et al.*, 2000; Zhang *et al.*, 2000). Studies of archaeal family Y polymerases, or DinB homologs, have tended to focus on the crenarchaea, particularly the Sulfolobus genus, and, as such, knowledge of the other phyla, including the euryarchaea, is relatively lacking (Lin *et al.*, 2010). Based on phylogenetic analyses, however, the Y family polymerases of archaea fall into one of five distinct clusters with crenarchaea and euryarchaea possessing distinctly different forms (Lin *et al.*, 2010).

1.6.5: Models of DNA replication and formation of replication forks

The process of chromosomal DNA replication shows significant similarities and differences between and within the three domains. Certain steps in the process are, however, universally conserved and will be discussed throughout the course of this chapter.

In all three domains the first step in chromosomal DNA replication involves the formation of an initiation complex at an AT-rich origin of replication (Aves, 2009). Initiation complex formation requires binding of initiator proteins, such as DnaA in *E. coli*, or origin recognition complex (ORC) and cell division cycle (CDC) proteins in archaea and eukaryotes. Through their interaction with, and recruitment of, other proteins, including the archaeal/eukaryotic GINS complex and MCM helicase, initiator proteins promote the unwinding of double-stranded DNA to form a replication bubble. As other elements bind to the newly exposed single-stranded DNA a replication fork is produced, as shown in figure 1.15. The main components of the replication fork include: the proliferating cell nuclear antigen (PCNA), replication factor C (RFC), replication protein A (RPA) and flap endonuclease 1

(FEN1). PCNA serves as a processivity clamp for DNA polymerases; it interacts with and recruits huge numbers of proteins involved in DNA replication and repair (Moldovan *et al.* 2007). RFC, also known as the clamp loader, is required for loading PCNA onto DNA in an ATP-dependent manner (Indiani and O'Donnell, 2006). RPA in archaea and eukaryotes, or single-stranded DNA binding protein (SSB) in *E. coli*, binds single-stranded DNA to prevent the formation of secondary structures and provide protection against nucleolytic attack (Oakley and Patrick, 2010; Meyer and Laine, 1990). FEN1 is involved in the processing of RNA flaps at the ends of archaeal and eukaryotic Okazaki fragments (Liu *et al.*, 2004). In bacteria the 5'-3' exonuclease (or flap endonuclease) domain of Pol I is responsible for processing Okazaki fragments and is contained within the C-terminus of the enzyme (Ogawa and Okazaki, 1984). Many bacteria, however, also possess a second FEN-encoding gene (Sayers, 1994).

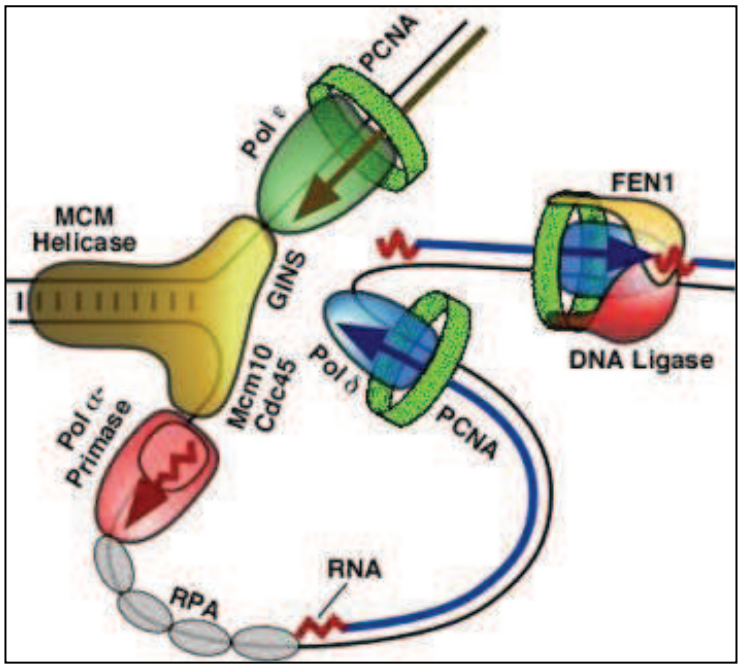


Figure 1.15: Components of the eukaryotic DNA replication fork (Garg and Burgers, 2005.

Critically, because polymerases can only synthesise DNA in the 5'-3' direction, replication occurs differently on the leading and lagging strands of replication forks. As previously stated, the first step in DNA replication involves synthesis of an RNA

primer by a primase (as described in chapter 1.6.3.4). The RNA primer is then elongated by Pol α in eukaryotes. In eukaryotes RFC displaces Pol α and initiates the switch to replication by the major replicases, Pol δ and ε (Yuzhakov *et al.*, 1999). Replication of the leading strand is performed by Pol δ and occurs in a continuous manner with Pol δ following closely behind the replication fork until adjacent forks converge. In euryarchaea it is thought that leading strand replication is performed by the family D polymerase (Pol D) (Henneke *et al.*, 2005). In all domains the lagging strand, however, must be replicated in a discontinuous manner with the polymerase extending away from the replication fork to produce fragments of 1000-2000 bases long in bacteria or 100-200 bases long in archaea and eukaryotes (Alberts *et al.*, 2002; Matsunaga *et al.*, 2003), each referred to as an Okazaki fragment. These chimeric DNA/RNA fragments then require further processing to remove the RNA primer. In bacteria this task is performed by the 5'-3' exonuclease domain of Pol I (Gutman and Minton, 1993). In archaea and eukaryotes, however, a specialised FEN1 enzyme performs the role (Lieber, 1997; Friedrich-Heineken and Hübscher, 2004). Okazaki fragment maturation, a process that occurs up to 50 million times per cell cycle in mammalian cells (Zheng and Shen, 2011), is then completed in eukaryotic cells through the combined actions of DNA polymerase δ, ligase 1 and a host of nucleases including RNAseH and DNA 2 (Zheng and Shen, 2011).

Due to the differing mechanisms involved, the process of chromosomal DNA replication is described as semi-discontinuous. In bacteria it appears that DNA polymerase III is responsible for the majority of chromosomal DNA replication, while DNA polymerase I is responsible for processing Okazaki fragments due to its flap endonuclease activity (O'Donnell, 2006). As previously stated, it is now widely accepted that DNA polymerase ε is responsible for leading strand replication in eukaryotes (Pursell *et al.*, 2007), while DNA polymerase δ is thought to replicate the lagging strand (Garg and Burgers, 2005; Nick McElhinny *et al.*, 2008). Pursell and colleagues demonstrated an inducible strand-specific bias in a>t and t>a transversion mutations by *in vivo* manipulation of the catalytic domain of *S. cerevisiae* Pol ε. At present the roles of archaeal polymerases in chromosomal DNA replication remain to be conclusively shown. Based on the respective strand displacement activities of euryarchaeal Pol B and Pol D, it appears likely that they may be responsible for leading and lagging strand replication, respectively, in a manner homologous to

eukaryotic DNA replication (Henneke *et al.*, 2005). As yet, however, there is no direct evidence to suggest that this is the case.

1.6.6: Biotechnological applications of DNA polymerases

1.6.6.1: Polymerase chain reaction (PCR)

PCR, first developed in 1983, describes the technique by which exponential amplification of a target DNA sequence is achieved through successive rounds of cycling (Bartlett and Stirling, 2003). The process requires a pair of primer oligodeoxynucleotides, which anneal at opposite ends and on opposite strands of the target DNA sequence. Replication is then promoted by addition of dNTPs and a processive DNA polymerase. After the first round of replication the newly synthesised double-stranded DNA must be denatured by heating to roughly 90-98 ºC to enable the excess of primers in the reaction to anneal to the newly-synthesised, denatured template DNA for the next round of replication. Originally Pol I, a polymerase of the mesophilic bacterium *E. coli*, was used in PCR (Saiki *et al.,* 1985); as such, the enzyme was denatured after each heat step and the reaction had to be supplemented with additional polymerase between successive rounds of replication. The application of thermostable polymerases in the PCR means that this is no longer the case (Saiki *et al.*, 1998)

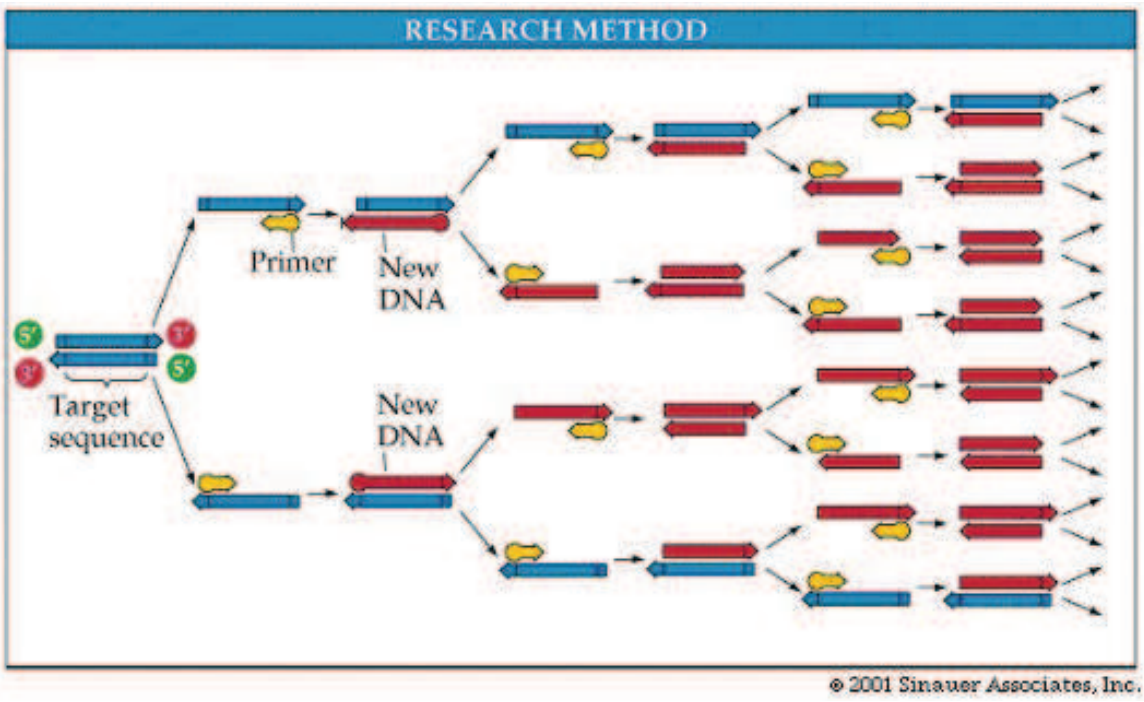


Figure 1.16: A schematic diagram demonstrating how exponential amplification of DNA is achieved by the PCR (Purves *et al.*, 2001)

The application of *Taq* DNA polymerase to PCR was a landmark in DNA research as its ability to withstand extreme temperatures meant that, compared to *E. coli* Pol I, it was exquisitely well suited to PCR (Saiki *et al.*, 1988). Repeated cycling at high temperatures does not result in rapid denaturation of *Taq* and, as such, reactions could for the first time be performed unintermittedly. The commercial appeal of such polymerases led Roche to pay US$300 million in 1991 for exclusive rights to market *Taq* and other thermostable polymerases. However, since then, controversy has surrounded the decision to award the patent on the grounds that application of native *Taq* polymerase to PCR does not represent a novel adaptation. As such, the patent has since been overturned in a succession of rulings (Service, 2001).

Due to the, albeit controversial, commercial success of *Taq* polymerase, considerable research on thermostable polymerases has since taken place. In particular, significant interest in archaeal polymerases exists for a number of reasons. Firstly, as previously mentioned, profound similarities can be seen in archaeal and eukaryotic systems, which is appealing from a research perspective. Secondly, the structural stability of thermophilic archaea is an essential property for polymerases in PCR. The extreme

environments that hyperthermophilic archaea inhabit make their polymerases ideally suited to successive rounds of heating and cooling. *Pyrococcus furiosus*, for example, possesses an optimal growth temperature of around 100 ºC (Fiala and Setter, 1986).

Since the discovery of *Taq* polymerase, numerous other thermostable DNA polymerases have now been identified and marketed for use in PCR, for example the family B polymerase of *P. furiosus* (Pfu-Pol B). Unlike *Taq* polymerase, Pfu-Pol B possesses a 3'-5' exonuclease domain and, as such, exhibits error rates roughly six fold lower than *Taq* (Cline *et al.*, 1996). Various technological developments have also been made to increase the commercial appeal of thermostable DNA polymerases. Chimeric fusion proteins of replicative polymerases and processivity clamps, such as Sso7d of *Sulfolobus solfataricus*, have been shown to significantly increase the processivity of the polymerase (Wang *et al.*, 2004). As such, these fusion polymerases are capable of efficiently amplifying much larger PCR products.

Since its initial discovery PCR has been applied to countless settings, ranging from clinical diagnostics and tissue typing to forensics and phylogenetic testing (Castle *et al.*, 2011; Olerup and Zetterquist, 1992; Willard *et al.*, 1998; Zhou *et al.*, 1998). Just a few of the research-based applications include quantification of DNA, generation of mutant libraries and construction of chimeric hybrid DNAs (Klein, 2002; Cirino *et al.*, 2003; Horton *et al.*, 1989).

1.6.6.2: DNA sequencing

DNA polymerases are also critical in most of the DNA sequencing technologies employed today. Sequencing typically involves amplification of a target sequence using a high fidelity DNA polymerase with a specific primer and a reaction mix comprising mainly dNTPs, but also a smaller population of labelled dideoxynucleoside triphosphates (ddNTPs) (Sanger *et al.*, 1977). Critically, ddNTPs lack a 3' hydroxyl group and are thus capable of being incorporated into growing DNA strands but not extended from. As such, incorporation of a ddNTP results in chain termination of the growing DNA strand. Amplification under these conditions therefore results in the production of a library of polynucleotides of varying length ranging from 1 to over 1000 bases (http://www.longtrace.com). Traditionally,

sequencing of a target DNA required four separate reactions; each using a different ^{32}P-labelled ddNTP to enable identification of the terminal base in a polynucleotide of known length. Since its inception, however, advancements have been made to improve the efficiency of sequencing reactions. The development of ddNTP-specific fluorescent labels, for example, means that a single sequencing reaction can now be performed in place of the four that were originally required (Prober *et al.*, 1987). Other adaptations, such as primer walking and shotgun sequencing, mean that whole genome sequencing is now relatively straightforward and achievable in remarkably short time frames (Fleischmann *et al.*, 1995).

1.6.6.3: Pharmaceutical interest in DNA polymerases

Polymerases in general also represent an appealing candidate for pharmaceutical compounds aimed at combating viral infections (Őberg, 2006). A significant number of nucleoside and pyrophosphate analogs, which target RNA and DNA polymerases, has been successfully employed in the treatment of viral diseases. Ribavirin, for example, is a prodrug, which is metabolised to form an RNA purine nucleoside analog. It has been shown to be effective against a variety of viral diseases, including influenza and hepatitis C (Sidwell *et al.*, 1979). Foscarnet, on the other hand, is a pyrophosphate analog that selectively inhibits the pyrophosphate-binding site of viral DNA polymerases, such as HIV reverse transcriptase (Oberg, 1989; Crumpacker, 1992).

Due to their fundamental role in cell proliferation DNA polymerases are also the target for a number of chemotherapeutic anti-cancer agents (Miura and Izuta, 2004). As with the antiviral compounds previously described, anti-cancer agents are often chain terminating nucleoside analogs. Examples of such compounds include fludarabine, cladribine, gemcitabine and cytarabine (Berdis, 2008). Fludarabine is a deoxyadenosine analog that contains arabinose in place of deoxyribose and is employed as a highly effective chain terminator (Huang *et al.*, 1990). Cytarabine, on the other hand, is unusual in that it inhibits both DNA replication *and* transcription (Datta *et al.*, 1992). Other analogs, such as decitabine, prevent methylation at sites of incorporation and thus interfere with gene regulation in infected cells (Brueckner *et al.*, 2007).

1.7: DNA deamination

Deamination simply describes the hydrolytic removal of an amine group from a molecule. The process can occur enzymatically, for example in the breakdown of amino acids in the human digestive tract to form ammonia. Spontaneous DNA deamination can also occur and affects bases that contain exocyclic amino groups. As with enzymatic deamination, spontaneous DNA deamination takes place in the form of a hydrolysis reaction and results in the production of ammonia. While cytosine, guanine and adenine are all capable of undergoing spontaneous deamination to form uracil, xanthine and hypoxanthine, respectively, the half-lives of these events differ considerably, as summarised in table 1.2. The frequency of each of these events is temperature-dependent, however, at biologically relevant temperatures spontaneous cytosine deamination occurs roughly 50-500 fold more frequently than that of either adenine or guanine. A schematic diagram depicting the spontaneous deamination of cytosine, adenine and guanine is shown in figure 1.17.

		25 °C	92 °C
Glyosidic cleavage	Cytosine	230 yrs	25 days
	Guanine	70 yrs	6 days
	Adenine	180 yrs	13 days
	Thymine	100 yrs	7 days
Deamination	Cytosine	120 yrs	20 days
	Guanine	60,000 yrs	6 yrs
	Adenine	20,000 yrs	3 yrs
Phosphodiester cleavage		31,000,000 yrs	3,500 yrs

Table 1.2: A summary of the half-lives of a number of common chemical bonds contained within normal double-stranded DNA at 25 ºC and 92 ºC. (Schroeder and Wolfenden, 2007, re-drawn by Emptage, 2009)

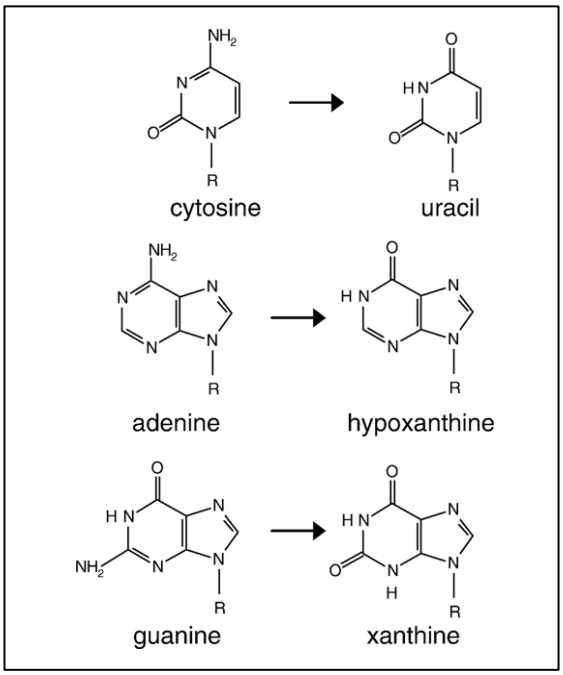


Figure 1.17: Structures of cytosine, adenine and guanine together with their deamination products: uracil, hypoxanthine and xanthine, respectively. Only bases that contain an exocyclic amino group are capable of undergoing deamination.

While all spontaneous DNA deamination has potentially detrimental consequences for living organisms, spontaneous cytosine deamination is of particular biological significance as it occurs considerably more frequently than that of either adenine or guanine. Due to the structural similarity of uracil and thymine residues, which differ by a single methyl group, uracil is completely compatible with Watson Crick base pairing (as shown in figure 1.18). Ordinarily, the additional methyl group of thymine protrudes into the major groove of DNA and does not affect the overall structure. Most polymerases therefore simply incorporate adenine opposite template-strand uracil (Takasawa *et al.*, 2004; Vaisman *et al.*, 2006). The net result of this is that 50 % of progeny end up containing a fixed c>t/g>a transition mutation at loci where replicative bypass of deaminated cytosine has occurred (Greagg *et al.*, 1999). Similarly, copying of hypoxanthine, which most closely resembles guanine, results in an a>g/t>c transition mutation in 50 % of progeny where the hypoxanthine residue has arisen from adenine deamination. In order to prevent such pro-mutagenic events, organisms from all three domains of life possess specialised repair enzymes that catalyse the removal of deaminated bases from DNA. Summaries of two such repair

enzymes, uracil DNA glycosylase (UDG) and endonuclease V, are provided in chapters 1.9 and 1.10, respectively.

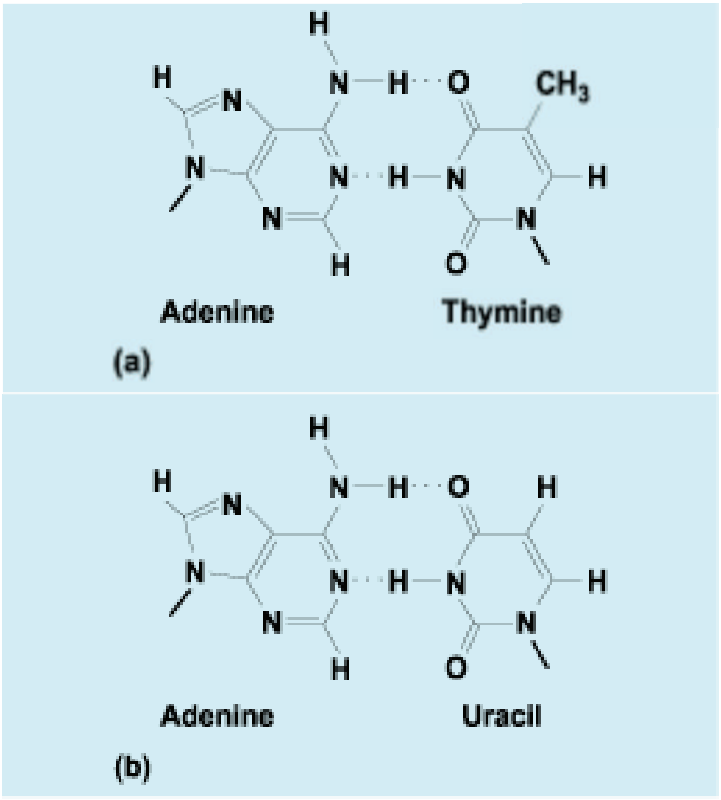


Figure 1.18: Comparison of **a)** an adenine:thymine base pair and **b)** an adenine:uracil base pair. (adapted from: http://www.accessscience.com/popup.aspx?figID=441200FG0070&id=441200&name=figure)

As well as being temperature-dependent, spontaneous DNA deamination can also be induced chemically through the use of nitrites and bisulfites (Shapiro and Pohl, 1968; Chen and Shaw, 1993). The efficacy of such treatments in studies of cultured organisms, however, is pH-dependent and largely specific to single-stranded DNA.

1.8: Deaminated base recognition by DNA polymerases

Archaeal DNA polymerases are unique in their ability to recognise and respond to deaminated bases, such as uracil and hypoxanthine. All archaeal family B polymerases tested to date possess a read ahead recognition ability, which enables them to detect template strand deaminated bases during DNA replication causing the polymerase to stall at a position 4 bases upstream from the lesion, as shown in figure 1.17 (Greagg *et al.*, 1999; Connolly, 2009).

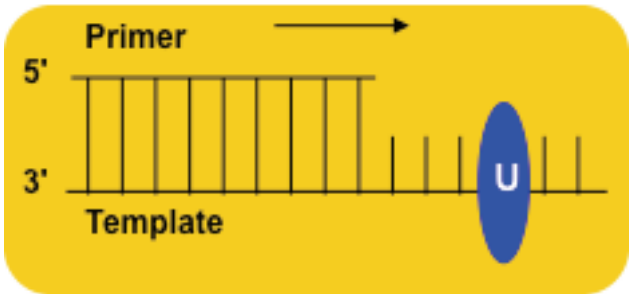


Figure 1.19: Position of archaeal Pol B stalling during replication upon encountering template strand uracil (unpublished diagram by B A Connolly, 2009).

A major source of interest in deaminated base recognition by archaeal family B polymerases stems from the ability of the enzyme to differentiate between bases that are structurally very similar and to do so with extreme efficiency; a feat which is all the more impressive since it is achieved while the polymerase replicates at remarkable speeds. Under optimal conditions the entire 1.9 Mb genome of *Pyrococcus furiosus* is replicated every 37 minutes (Fiala and Stetter, 1986). Discrimination between undamaged and deaminated bases must therefore be achieved without hindering this process. Reference to figures 1.17 and 1.18 reveals how similar cytosine, adenine and guanine are to their deamination products.

Structural studies of deaminated base recognition in archaeal family B polymerases revealed that a "uracil-binding pocket", located in the N-terminus of the enzyme, is responsible for this activity (Fogg *et al.*, 2002; Firbank *et al.*, 2008). The uracil-binding pocket is capable of trapping deaminated bases with extreme efficiency, thereby preventing replicative bypass of the lesion. In fact Pfu-Pol B has been shown

to interact with single-stranded DNAs containing uracil 200-300 times more strongly than it does with equivalent substrates that contain only the four canonical bases (Shuttleworth *et al.*, 2004; Gill *et al.*, 2007). More recently it was shown that the uracil-binding pocket appears to rely on a combination of steric compatibility and hydrogen bonding (Killelea *et al.*, 2010). Steric complementarity between the pocket and the C5 atom of uracil (or the C2 atom of hypoxanthine) appears to be particularly crucial for recognition. While hydrogen bonding between the O^2, N^3 and O^4 atoms of uracil (or the equivalent O^6, N^1 and N^3 atoms of hypoxanthine) is also required for specific detection and trapping of deaminated bases (Gill *et al.*, 2007). Mutation of the 93rd residue of Pfu-Pol B from valine to glutamine has been shown to completely eliminate the enzyme's ability to detect and respond to deaminated bases (Fogg *et al.*, 2002).

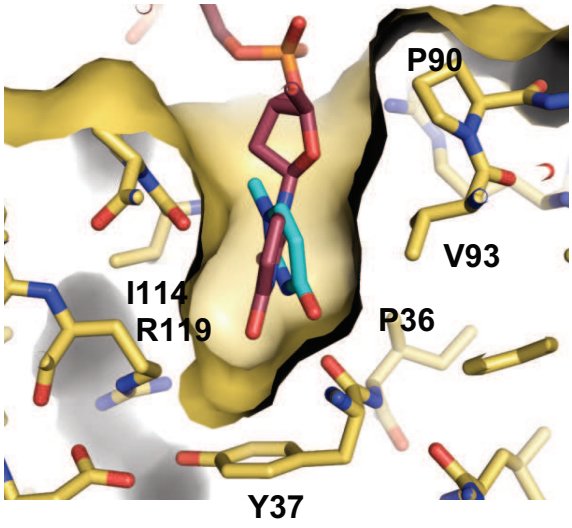


Figure 1.20: A crystal structure, solved from X-ray diffraction patterns, of the uracil-binding pocket of *Thermococcus gorgonarius* family B DNA polymerase with a uracil residue tightly bound. The six key amino acid residues involved in formation of the pocket are labelled in black (Firbank *et al.*, 2007).

It is thought that archaea may require additional protection against base deamination, compared to bacteria and eukaryotes, due to their propensity for physically and geochemically extreme environments. *P. furiosus*, for example, possesses an optimal

growth temperature of around 100 ºC (Fiala and Stetter, 1986) while members of the *Pyrodictium* genus are capable of surviving at temperatures as high as 110 ºC (Stetter, 1982). At such extreme temperatures, as well as experiencing increased rates of spontaneous DNA deamination, it is thought that hyperthermophilic organisms may also transiently produce more single-stranded chromosomal DNA, which is itself more susceptible to spontaneous deamination (Frederico *et al.*, 1990).

While replicative stalling by archaeal family B polymerases may constitute the first step in a novel and as-yet uncharacterised repair pathway (Greagg *et al.*, 1999), currently the events that occur during archaeal DNA replication post-stalling remain unknown. Since strand termination and cell death would seem counter-intuitive for a single celled organism, it seems more likely that stalling serves to promote the formation of complexes, possibly involving repair enzymes such as UDG or EndoV. Unfortunately, as yet, no evidence exists to suggest that this is the case. It has also been suggested that polymerase stalling on exposed single-stranded DNA may initiate a recombinational gap repair process (Greagg *et al.*, 1999). Certainly in eukaryotes single-stranded DNA may serve as the primary signal for DNA repair (Garvik *et al.*, 1995).

Interestingly, the family B polymerases of archaea are capable of recognising and responding to uracil and hypoxanthine, but not xanthine (Gill *et al.*, 2007). From a biological perspective, recognition of hypoxanthine and uracil is favourable due to the greater frequency of spontaneous adenine, and particularly cytosine, deamination (Schroeder and Wolfenden, 2007). However, the exact reasons for the absence of xanthine recognition remain unclear. It may be that null recognition of xanthine is a corollary of the need to distinguish between guanine and hypoxanthine. This distinction relies heavily on the steric exclusion of the exocyclic $-NH_2$ group of guanine and may therefore result in exclusion of the equivalent =O side chain of xanthine. However, since the catalytic site of archaeal family B polymerases is unable to copy xanthine (Gill *et al.*, 2007), a separate stalling response induced by the pocket may be unnecessary to prevent replicative bypass.

Sequence homology between eukaryotic and archaeal DNA polymerases suggests that eukaryotic family B polymerases may also possess a uracil-recognition domain.

However, biochemical analyses have shown that eukaryotic family B polymerases do not recognise deaminated bases and any vestigial uracil-binding pocket is non-functional (Tahirov *et al.*, 2009; Wardle *et al.*, 2008). It seems likely that, evolutionarily speaking, as global climate shifted and eukaryotes gravitated towards more mesophilic environments, the uracil recognition domain of polymerases became increasingly redundant. Both archaea and eukaryotes possess specialised uracil and hypoxanthine repair machinery, enabling the efficient excision of deaminated bases from double-stranded DNA. It would therefore appear that the higher frequency of base deamination associated with a hyperthermophilic lifestyle has provided sufficient evolutionary pressure to maintain the functionality of the uracil-recognition domain in archaeal DNA polymerases, but not in the predominantly mesophilic Eukarya. Interestingly, however, comparisons of mesophilic archaea and hyperthermophilic bacteria appear to show no obvious correlation between a hyperthermophilic lifestyle and possession of polymerases that are capable of deaminated base recognition (Wardle *et al.*, 2008). The DNA polymerase of the hyperthermophilic bacterium *Taq*, for example, is incapable of recognising deaminated bases, while the family B polymerase of *Methanosarcina acetivorans*, a mesophilic archaeon, appears to have retained the ability. There is also no evidence to suggest that, evolutionarily speaking, bacterial replicative polymerases ever possessed a uracil recognition domain. This fundamental difference is indicative of the early evolutionary divergence of bacteria and archaea.

One of the major drawbacks of using hyperthermophilic archaeal DNA polymerases in PCR-based applications is the fact that they *do* stall DNA replication upon encountering deaminated bases, resulting in truncated products where such DNA damage exists. The common practice of using dUTP in place of dTTP in PCRs to prevent cross-contamination is therefore incompatible with wild type archaeal polymerases (Longo *et al.*, 1990). The identification of uracil-insensitive mutant variants, such as Pfu-V93Q (Fogg *et al.*, 2002), therefore provides useful biotechnological applications.

1.9: Uracil DNA glycosylases (UDGs)

UDGs are found in nearly all cellular organisms, as well as pox and herpes viruses (Aravind and Koonin, 2000). They catalyse the efficient removal of uracil from both single and double-stranded DNA to generate an AP site. They are also capable of acting upon various oxidised cytosine derivatives (Dizdaroglu, 1996). UDGs preferentially locate and excise uracil from U/A and U/G base pairs, which result from dUMP incorporation and cytosine deamination, respectively. From a biological perspective, U/A base pairs pose a less serious threat to living cells, as they are non-mutagenic. The presence of uracil, however, still represents a major source of AP sites and may well be cytotoxic (Hagen *et al.*, 2006). Uracil causes disruption of DNA interactions and specific binding of proteins (Zharkov *et al.*, 2010). U/G base pairs, on the other hand, are defined as pro-mutagenic and, if left uncorrected, lead to c>t/g>a transition mutations in 50 % of progeny.

The first step in the UDG repair pathway involves detection of a spontaneously "flipped" uracil residue. Spontaneous flipping occurs due to the dynamic motions of DNA base pairs and the relatively high spontaneous opening rates of U/A and especially U/G base pairs in double-stranded DNA (Friedman and Stivers, 2010). With the flipped uracil bound to the catalytic site, the residue is then cleaved. At this stage, due to the combined kinetic and thermodynamic barriers selecting against it, UDG excises thymine residues 10^8 times more slowly than uracil (Friedman and Stivers, 2010). The enzyme is therefore capable of discriminating between largely isosteric bases with remarkable efficiency. The cleavage reaction takes place by hydrolysis of the *N*-glycosidic bond between uracil and deoxyribose to generate an AP site. In double-stranded DNA this initiates a base excision repair pathway, in which an AP endonuclease recognises the lesion and breaks the phosphodiester backbone 5' to the AP site. Full repair of the lesion then takes place by incorporation of deoxycytidine monophosphate, which is performed by a polymerase. Finally, the nicked DNA is rejoined by the action of DNA ligase (Yonekura *et al.*, 2009).

Mechanistically speaking, UDGs are capable of cleaving uracil from both single- and double-stranded DNA with a slight preference for the former (Lindahl *et al.*, 1977).

This breadth of substrates is notable in that it necessitates a microscopic dissociation event by the enzyme to facilitate switching between the two strands (Friedman and Stivers, 2010). In order to locate uracil residues, it has recently been shown that the UDG of *E. coli* predominantly operates using a hopping mechanism to move between distant regions of DNA, later followed by a short range sliding mechanism, which covers regions of approximately 10 bp (Porecha and Stivers, 2008). Upon associating with the chromosome, UDG either adopts an open form that interacts weakly with DNA and is competent for translocation by sliding, or a closed form that is capable of interrogating base pairs (Friedman and Stivers, 2010).

1.10: Endonuclease V

EndoV is another repair enzyme that is responsible for cleaving deaminated bases from DNA and is found in all 3 domains of life (Dalhus *et al.*, 2009). It is primarily involved in the excision of hypoxanthine, the deamination product of deoxyadenosine, but is also capable of acting upon mismatched bases, as well as the other two products of DNA deamination, uracil and xanthine (Schouten and Weiss, 1999; Feng *et al.*, 2005). Uniquely, the enzyme functions by catalysing the cleavage of the 2nd phosphodiester bond 3' to the targeted residue, using a Mg^{2+}-dependent mechanism to generate a nicked product (Yao *et al.*, 1994). The downstream events that occur after nicking are currently unknown, however, given the high affinity of EndoV for nicked substrates and the presence of a 3'-5' exonuclease domain (Mi *et al.*, 2011), it seems plausible that the enzyme may generate a DNA repair patch spanning 2-3 nucleotides either side of the targeted residue (Weiss, 2008).

1.11: Transciription-coupled repair (TCR)

Although first reported in a eukaryote (Bohr *et al.*, 1985), the phenomenon of transcription-coupled repair has largely been characterised through studies of *E. coli*

using UV-induced DNA damage (Mellon and Hannawalt, 1989). Despite considerable research occurring in the interim years, both on eukaryotic and bacterial TCR pathways, as yet there is no evidence to suggest the existence of a comparable system in the Archaea (Dorazi *et al.*, 2007; Romano *et al.*, 2007). The observation that active genes undergo significantly higher rates of repair than inactive regions of bacterial and eukaryotic genomes led to research which demonstrated that the difference in repair frequency does not simply result from increased accessibility of transcribed regions, but instead results from a more sophisticated mechanism. This conclusion arose from the observation that the transcribed strand undergoes a significantly higher rate of repair than the untranscribed strand of the gene in question (Mellon and Hannawalt, 1989).

The current, putative model for TCR in bacteria suggests that the increased rate of repair in transcribed strands results from a complex series of events involving the RNA polymerase and a number of interacting partners (Savery, 2011). Upon encountering DNA lesions, RNA Pol stalls, which promotes binding of the ATP-dependent DNA translocase, Mfd. Mfd binding then causes RNA Pol to dissociate, exposing the lesion and enabling binding of the nucleotide excision repair (NER) proteins. Through a complex interplay involving UvrA, UvrB, UvrC and UvrD, the NER pathway enables unwinding of the DNA, removal of the lesion and fill up of the repaired strand. Since the development of this model, however, evidence has emerged, which suggests the existence of at least one alternative, Mfd-independent pathway (Savery, 2011).

Despite the majority of early work on TCR occurring in bacteria, considerable interest in the eukaryotic system exists due to its relevance to human genetic diseases such as Cockayne syndrome. Significantly, in contrast to bacterial TCR, it is known that the base excision repair (BER) pathway is also involved in the eukaryotic process, as well as a much larger number of interacting partners. So far, human proteins including CSA, CSB, XAB2, TFIIH and XPG have all been shown to play a major role in TCR repair pathways (Svejstrup, 2002). Despite the greater number of identified partners, significant doubts remain regarding the details of TCR in eukaryotes, particularly with regard to the step involving displacement of RNA Pol II. Interestingly, however, it *is* know that the eukaryotic process involves hyperphosphorylation and ubiquitylation of

the RNA polymerase, and thus rapid degradation of the enzyme in response to lesion encounter (Svejstrup, 2002).

1.12: Aims

The principal aim of this PhD is to further elucidate the mechanisms responsible for deaminated base recognition and response in archaeal family B DNA polymerases; particularly in terms of the events that occur post-stalling. Specifically, the role of 3'-5' exonucleolytic activity during the deaminated base response will be studied. The role of DNA unwinding by archaeal Pol B in response to template strand uracil will also be probed.

Another of the key aims of this study is to examine the biological significance of deaminated base recognition in polymerases and whether or not this ability significantly contributes to maintaining genomic stability. This issue is of particular interest considering the presence of alternative repair pathways for deaminated bases. A range of genetic techniques will be employed using the model species, *Thermococcus kodakarensis*, to achieve this end.

The other main focus of this project will be the somewhat enigmatic family D DNA polymerase, which is found exclusively in members of the Euryarchaea and several emergent archaeal phyla. Biochemical studies aimed at identifying and characterising a uracil recognition ability in family D DNA polymerases will be described in chapter 6.

Chapter 2

Materials and Methods

2.1: Oligodeoxynucleotide design, synthesis, purification and annealing

2.1.1: Oligodeoxynucleotide design and synthesis

All oligodeoxynucleotides were designed using Clone Manager Professional Suite version 8.0 (Scientific & Educational Software, 600 Pinner Weald Way Ste 202, Cary NC 27513, USA) and synthesised on an Applied Biosystems 392 DNA/RNA synthesizer using columns and reagents from Proligo (6200 Lookout Rd Boulder, CO 80301, USA). Following synthesis, oligodeoxynucleotides were cleaved from their glass bead supports by incubation at 55 ºC in 35 % ammonia for 4 - 15 hours. The ammonia was then removed by evaporation using a Savant Speed Vac SC100 and the glass bead supports separated from solution by filtration through a 22 μm Millipore Millex syringe-driven filter unit.

2.1.2: Oligodeoxynucleotide purification

Purification of oligodeoxynucleotides was performed by reverse phase high-pressure liquid chromatography (HPLC). Purification was performed at 60 ºC using an Apex C18 octadecylsilyl 0.5 micron column (Jones Chromatography, Llanbradach, Wales). Two buffers were employed: buffer A (5 % acetonitrile; 0.6 % acetic acid, adjusted to pH 6.5 using triethylamine) and buffer B (65 % acetonitrile; 0.6 % acetic acid, adjusted to pH 6.5 using triethylamine). The column was pre-equilibrated at 1 ml/min. for 20 minutes. Sample were loaded at 1 ml/min. and eluted using a 25 ml linear gradient of buffer A to B again at 1 ml/min. The buffer gradient was varied depending on the purification handle employed, as outlined in table 2.1.

Purification handle	Gradient (buffer B)
None	0-20 %
Dimethoxytrityl (DMT) group	25-75 %
Cy5	10-65%
Hexachlorofluorescein	10-50%

Table 2.1: Buffer gradients used for HPLC purification of oligodeoxynucleotides possessing different purification handles.

Eluted oligodeoxynucleotides were detected by absorbance at 260 nm. In the case of unlabelled oligodeoxynucleotides requiring post-synthesis purification, the 5' dimethoxytrityl (DMT) group was removed following purification by incubation at room temperature in 80 % Analar acetic acid for 1 hour. All traces of acetic acid were then removed by evaporation in a rotary evaporator followed by resuspension in 20 ml of nanopure water. The volume of water was then reduced to ≈0.5 ml by evaporation in a rotary evaporator. This wash step was repeated twice more before the purified oligodeoxynucleotide was resuspended in a total volume of 1 ml of nanopure water and stored at -20 ºC.

2.1.3: Determination of oligodeoxynucleotide concentration

Primer concentrations were calculated using the Beer-Lambert Law:

$$C = A_{260}/\varepsilon.l$$

Where C = the mM concentration of oligodeoxynucleotide; A_{260} = the absorbance of light of wavelength 260 nm. ε = the extinction coefficient of the oligodeoxynucleotide at 260 nm ($mM^{-1}cm^{-1}$), calculated by adding the individual extinction coefficients of each base (as outlined in table 2.2). l = the path length of the quartz cuvette.

Deoxynucleotide	Extinction coefficient ($mM^{-1}cm^{-1}$)
dA	14.7
dC	6.1
dG	11.8
dT	8.7
dU	10.0
dH	7.5
2AP	1.0
2-pyrimidinone	6.5*
Cy5	10.0
Hex	31.6
Fluorescein	38.8

Table 2.2: Extinction coefficients of individual base components of oligodeoxynucleotides at 260 nm in single-stranded context. (http://www.glenresearch.com//Reference/Extinctions.html)

* Gildea and McLaughlin (1989)

2.1.4: Annealing of oligodeoxynucleotide duplexes

After determining the molar concentration of purified oligodeoxynucleotides, DNA duplexes were annealed at a molar ratio of between 1:1 and 1:2, depending on the application, with the unlabelled strand always in excess. In the case of 2-aminopurine fluorescence studies, primer extensions and exonuclease assays, a two-fold excess of unlabelled template was used. In the case of fluorescence anisotropy studies, a 1.3 fold excess of unlabelled oligodeoxynucleotide was used. In the case of Pfu-Pol D strand separation assays, a 1:1 molar ratio was used. Duplexes were annealed in 10 mM Hepes-NaOH (pH 7.5), 100 mM NaCl and 1 mM EDTA by incubation at 95 ºC for 5 minutes followed by cooling by 0.5 ºC per minute to room temperature using either a heat block or thermocycler. Annealed DNA duplexes were stored at -20 ºC then thawed at room temperature as required.

2.2: PCR reactions

Standard 50 µl scale PCR reactions were performed using: 1-10 ng/µl of template DNA; 200 µM dATP; 200 µM dTTP; 200 µM dCTP; 200 µM dGTP; 1 µM of each primer; 1.25 U of *Taq* (Fermentas) or 2 U of Velocity polymerase (Bioline) in 1x manufacturer's reaction buffer, supplemented with 1-3 mM $MgCl_2$ where optimisation was required, and topped up to 50 µl volume with nuclease-free water.

Thermocycling was carried out using either a PCR Sprint Thermo Hybaid thermocycler or a Biometra TPersonal Combi thermocycler using the following cycling conditions:

95 ºC for 2 minutes	x 1
95 ºC for 35 s T_m-5 for 35 s 72 ºC for 1 min./kb	x 30
72 ºC for 3 min./kb	x 1

(Where T_m = melting temperature of primers)

All primers used in this study for PCR and sequencing reactions are described in table 2.3 below.

Name	Use	Sequence
Pfu-Pol B 500FP	Sequencing	5′-GCAGATGAAAATGAAGCA-3′
DP1 H445A FP	Site-directed mutagenesis	5′-CAGGAAACgcCGATGCTGCTAGGCAAGCTATTC-3′
DP1 H445A RP	Site-directed mutagenesis	5′-CTAGCAGCATCGgcGTTTCCTGGGGCAATGAAC-3′
DP1 700FP	Sequencing	5′-GGAAAATACTAAGGGAAAATC-3′
DP2 C953A FP	Site-directed mutagenesis	5′-GCGGCCAAGAGAAGGAACgcTGATGGAG-3′
DP2 C953A RP	Site-directed mutagenesis	5′-CACTATCCTCATCTCCATCAgcGTTCCTTC-3′
DP2 2100FP	Sequencing	5′-CCGAAGTGTGGCCATGTAGG-3′
TK0001 25FP	Cloning (PaeI)	5′-GTTTCTTGCATGCAATGAGTGTTTGGAACAGTATCAATCCGTCCTTTC-3′
TK0001 2500RP	Cloning (PstI)	5′-GTTTCTTCTGCAGTTGGCGAGGTTTTCATCCCCAAAGC-3′
TK2303 25FP	Cloning (EcoRI)	5′-GTTTCTTGAATTCCATCAAGAGCGCCAGCGAGGCTGC-3′
TK2303 2000FP	Cloning (Acc65I)	5′-GTTTCTTGGTACCAAAGTTCAACTGCCTTCTCAAGGTAGCCAAG
TK D215A FP	Site-directed mutagenesis	5′-GGCGAAGgCGAAGTTGTCGCCGTTGTAG-3′
TK D215A RP	Site-directed mutagenesis	5′-GACAACTTCGcCTTCGCCTATCTG-3′
TK V93Q FP	Site-directed mutagenesis	5′-TTATCGCTGGctgGTCCTGCGGATGAGTAAAG-3′
TK V93Q RP	Site-directed mutagenesis	5′-CCGCAGGACcagCCAGCGATAAGGGACAAG-3′
TK2303 20FP	Sequencing	5′-CAGCGAGGCTGCCATAATGATCC-3′
TK2303 200RP	Sequencing	5′-CCACTAACCACAGCCCTGAAC-3′
TK2303 700FP	Sequencing	5′-TAAGCCGGTCGTAGGAGTAGTC-3′
TK2303 1400FP	Sequencing	5′-GCCACTTCTCGATCTCGAAGAC-3′
TrpE 200FP	Sequencing	5′-GAGCTCTTAAAGGGCTCATGG-3′
TK0001 20FP	Sequencing	5′-ACAGTATCAATCCGTCCTTTCC-3′
TK0001 200RP	Sequencing	5′-ACAGGGTTAGGAGGAGTATTG-3′
TK0001 700FP	Sequencing	5′-GCTTATGATTGGGCGTACAC-3′
TK0001 1400FP	Sequencing	5′-CCGAGGGCGAAGTTTATTCC-3′
TK0001 2100FP	Sequencing	5′-CCATCCTCGGTTATGTAGTC-3′
TK0001 2300RP	Diagnostic PCR/sequencing	5′-GCGAGACGTTGTGGGTGATGATGATTGAG-3′
TK2303 1300FP	Diagnostic PCR/sequencing	5′-GGTCGGCGGCAAGGAGCAGCTCGCCATCG-3′
TK0001 seq	Diagnostic PCR/sequencing	5′-CGAAGACGGCTTCATAAACGGCC-3′
TrpE FP	Diagnostic PCR	5′-CGGCCGATATGTAGGTGAATCTGGCCTTCCTGGAGTCCTTCTCGGCAGAG-3′
TrpE RP	Diagnostic PCR	5′-GCGATGGAGATCATAGACGAGCTGGAGAGGAGCAGGAGAAAGGTCTACG-3′
TK0664 FP	Amplification/Sequencing	5′-CGTCGATGCTCTCGCTGTGCATTTC-3′
TK0664 RP	Amplification/Sequencing	5′-AGCCCCAGAACACAATAGTCACATC-3′

Table 2.3: A summary of all primer oligodeoxynucleotides used in this study for PCR and sequencing reactions. In the case of cloning primers, "pigtails" are highlighted in blue and restriction sites in red; the name of the restriction endonuclease for use with each cloning primer is also shown in the "Use" column. In the case of site-directed mutagenesis primers, mutated regions are highlighted in green.

2.3: Agarose gel electrophoresis, DNA extraction and DNA quantification

2.3.1: Agarose gel electrophoresis

PCR products were analysed by agarose gel electrophoresis in 1x TBE buffer using 1x TBE/1 % agarose gels supplemented with 1µg/ml of ethidium bromide. Outer lanes were loaded with 500 ng of Fermentas GeneRuler 1 kb or 100 bp DNA ladder, depending on the product size. Samples were supplemented with 6.25 % glycerol, 0.625 % SDS, 0.1 % bromophenol blue and 0.1 % xylene cyanol. Gels were run at 100 V, 100 mA, 10 W until suitable resolution was achieved and the products visualised using a UV transilluminator.

2.3.2: Extraction of DNA bands

Where PCR products were required for downstream applications, DNA bands were manually excised from the agarose gel using a surgical blade following resolution by electrophoresis. The excised gel was then weighed and subjected to DNA extraction using either a QIAquick gel extraction kit (Qiagen) or a Zymoclean gel DNA recovery kit (Zymo Research) according to the manufacturer's guidelines.

2.3.3: Quantification of purified DNA

Concentration and purity of DNA samples were estimated by measurement of UV absorbance spectra. The A_{260} reading was used to estimate DNA concentration using the Beer-Lambert Law previously described, however, the extinction coefficient for nucleic acids was assumed to be 50 ng/µl (1 cm pathlength) for double-stranded DNA and 33 ng/µl (1 cm pathlength) for single-stranded DNA. The A_{260}:A_{230} ratio was used as an indication of purity with a ratio of 1.6 or above considered to be satisfactory.

2.4: Plasmid design and construct preparation

2.4.1: Plasmid and construct design

All DNA plasmids and constructs produced during the course of this PhD were designed using Clone Manager Professional Suite 8.0 (Scientific & Educational Software, 600 Pinner Weald Way Ste 202, Cary NC 27513, USA), NEB cutter V2.0 (New England Biolabs, 240 County Road, Ipswich, MA 01938-2723, USA) and PlasMapper Version 2.0 (Dong *et al.*, 2004).

2.4.2: Restriction digests

Restriction digests were all performed using Fermentas enzymes according to the manufacturer's guidelines.

2.4.3: Vector dephosphorylation

Vector dephosphorylation was performed when cloning to prevent re-circularisation of the digested vector. Dephosphorylation was carried out during the last 15 minutes of restriction digests at 37 ºC by adding 1/10 volume of 10x Antarctic Phosphatase Reaction Buffer (NEB) and 1 µl of Antarctic Phosphatase (NEB) per 1-5 µg of DNA. The phosphatase was then heat-inactivated at 65 ºC for 5 minutes before performing DNA purification and proceeding to ligation (as described below).

2.4.4: DNA purification

Extraneous enzyme, salt and buffer were removed from DNA samples after performing PCRs, restriction digests and gel extractions by performing QIAquick PCR purifications (Qiagen) according to the manufacturer's protocol.

2.4.5: Ligations

Ligations were carried out using T4 DNA ligase from Fermentas, according to the manufacturer's guidelines.

2.4.6: Preparation of chemically competent *Escherichia coli* by calcium chloride treatment

Competent bacterial cells were produced by inoculating 50 ml of LB medium with 0.5 ml of saturated monoclonal starter culture. The LB medium was supplemented with 30 ng/µl of chloramphenicol (CAM) in the case of BL21-Codon Plus (DE3)-RIL or BL21 (DE3) pLysS cells. After addition of the inoculum, cultures were incubated at 37 ºC with shaking (150 rpm) for roughly 2 hours until the OD_{600} reached 0.4-0.6. Cells were then harvested by centrifugation at 4000 g, 4 ºC for 10 minutes, followed by removal of the supernatant and resuspension in 10 ml of ice cold 100 mM $MgCl_2$. Cells were harvested once more by centrifugation at 4000 g, 4 ºC for 10 minutes, then resuspended in 2 ml of ice cold 100 mM $CaCl_2$ at which point the cells were ready to use. For long-term storage glycerol was added to a final concentration of 20 % (v/v) and the cells dispensed in 50-100 µl aliquots before storage at -80 ºC.

2.4.7: Transformation of chemically competent *E. coli*

Transformation of competent cells was performed by adding 1-10 ng of plasmid DNA per µl of $CaCl_2$-treated cells and incubating on ice for 30 minutes. Cells were heat-

shocked at 42 ºC for 30 s and put back on ice for 2 minutes. 5-10 µl of LB was added per µl of cells and the cell suspension incubated at 37 ºC for 1 hour. After 1 hour 50-100 µl of transformed cell suspension was plated on an LB-agar plate containing appropriate antibiotic(s), as described in chapter 2.4.8, and incubated at 37 ºC for 12-16 hours or until single colonies appeared.

2.4.8: Transformant selection and plasmid isolation

Bacterial transformants were selected by growth on LB-agar plates at 37 ºC using 1% agar, 1x LB and antibiotics at the following concentrations: ampicillin at 100 ng/µl; carbenicillin at 100 ng/µl; CAM at 34 ng/µl; and kanamycin at 50 ng/µl. After 12-16 hours single colonies were picked and grown in 5 ml of LB, containing appropriate antibiotic(s), at 37 ºC for 12-16 hours. Cells were then harvested by centrifugation and subjected to QIAprep Spin Miniprep (Qiagen) plasmid isolation, according to the manufacturer's protocol.

2.4.9: Site-directed mutagenesis

Site-directed mutageneses were carried out using primers, which were designed in Clone Manager Professional Suite 8.0, together with a QuikChange mutagenesis kit (Stratagene). PCR reactions were carried out using Velocity DNA polymerase (Bioline) and the QuikChange protocol performed as described by the manufacturer. After isolation, mutagenised plasmids were sent for full sequencing to confirm that the desired mutation had been produced and that there were no extraneous mutations present.

2.4.10: Preparation of plasmids for the production of M247A and Y261A mutant variants of Pfu-Pol B

Plasmids used for the expression of Pfu-Pol B loop mutants were based on the pET17b-Pfu-Pol B wild type (WT) and D215A vectors originally produced by Evans *et al.* (2000). M247A and Y261A mutations were introduced to the Pfu-Pol B gene using the Pol B M247A FP and RP primers and the Y261A FP and RP primers (shown in table 2.3), respectively. The mutations were introduced on their own and in combination with one another in both WT and D215A backgrounds to produce six new variants of Pfu-Pol B. The D215A background was used to prevent 3'-5' exonucleolytic degradation of oligodeoxynucleotides during fluorescence analyses. While the WT background was used to measure the effect of these mutations on the rates of exonucleolysis by Pfu-Pol B in response to deaminated bases.

2.4.11: Preparation of plasmids for the manipulation of *Thermococcus kodakarensis*

All four plasmids used in the manipulation of *T. kod* were based on pUMT2, which was supplied by Dr Tom Santangelo of Ohio State University, Columbus, USA. Two pieces of DNA were amplified from a contiguous region of *T. kod* KW128 genomic DNA and cloned either side of the pUMT2 TrpE marker. The first piece, a 2.5 kb region encoding the promoter and N-terminus of TK0001, was amplified using the TK0001 FP and TK0001 RP primers shown in table 2.3 and cloned upstream of the TrpE marker using the PstI and PaeI (SphI) restriction sites of pUMT2. The second piece, a 2 kb amplicon encoding TK2305, TK2304 and the C-terminus of TK2303, was amplified using the TK2303 FP and TK2303 RP primers shown in table 2.3 and cloned downstream of the TrpE marker using the EcoRI and KpnI restriction sites of pUMT2. The resulting plasmid, named pTR1, is shown in figure 2.1 below.

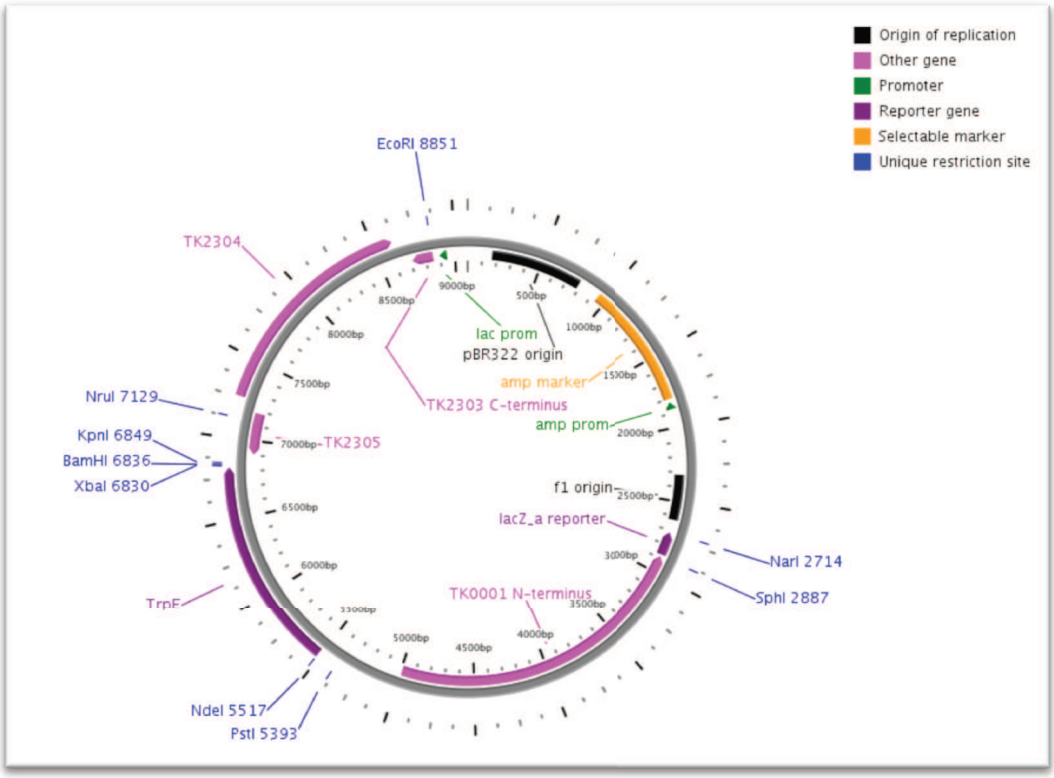


Figure 2.1: Plasmid map of pTR1, which was used for the manipulation of *T. kod* strain KW128 by promotion of allelic exchange. This image was generated using PlasMapper Version 2.0.

After confirming that the plasmid contained the desired cloned regions by performing diagnostic PCRs and full sequencing, D215A and V93Q mutations were introduced to the TK0001 gene using the TK D215A FP and RP primers and TKV93Q FP and RP primers, shown in table 2.3, respectively. The mutations were introduced individually and in combination with one another to produce three additional plasmids (termed pTR2-pTR4). Full sequencing confirmed that pTR2 contained the D215A mutation, pTR3 contained the V93Q mutation and pTR4 contained both the D215A and V93Q mutations.

2.4.12: Plasmids used for expression and purification of Pfu-Pol D

The original plasmids used for the expression of Pfu-Pol D during this PhD were generous gifts from Professor Yoshi Ishino of Kyushu University, Japan. The following plasmids were provided to the laboratory in 2009:

pET22b-PfuDP1 (non tagged)
pET28a-PfuDP1 (His tagged)
pET22b-PfuDP2 (non tagged)

PfuDP1 describes the small subunit, upon which the 3'-5' exonuclease domain resides and PfuDP2 described the large subunit, which is responsible for polymerase activity. Based on the work of Gueguen *et al.* (2001), we introduced the H445A mutation to the pET28a-PfuDP1 (His tagged) plasmid by QuikChange mutagenesis. This mutation, which corresponds to the H454 residue of Pab-Pol D, has previously been shown to completely eliminate 3'-5' exonucleolytic activity. Mutagenesis was performed using the DP1 H445A FP and RP primers shown in table 2.3. Full sequencing confirmed the presence of the desired mutation and that no extraneous mutations had been produced during the process. Exonuclease assay were used to confirm that the 3'-5' exonuclease activity had been completely eliminated, resulting in an exo- variant of Pfu-Pol D that was suitable for use in fluorescence anisotropy binding assays and single turnover primer extension assays.

Based on sequence analyses that were performed by our group, we also decided to introduce a C953A mutation to the pET22b-PfuDP2 subunit (non tagged) by QuikChange mutagenesis using the DP2 C953A FP and RP primers shown in table 2.3. Again, full sequencing confirmed the presence of the desired mutation and the absence of any extraneous mutations.

2.5: Protein expression, purification and analysis

2.5.1: Expression of Pfu-Pol B and Pfu-PCNA

WT and mutant versions of pET17b-Pfu Pol B (described in 2.4.10) were used for expression of Pfu-Pol B. The required plasmid was transformed into chemically-competent BL21 (DE3) pLysS cells (Stratagene) and selected for on LBagar-$carb_{100}$-CAM_{34} plates. A single colony was used to inoculate a 50 ml starter culture, which was grown to saturation overnight under selective conditions before being used to inoculate 500 ml of LB-$carb_{100}$-CAM_{34} using 0.5 % (v/v) inoculum. The large-scale culture was grown at 37 ºC, 150 rpm, for roughly 1-2 hours (until OD_{600} = 0.6), then induced with 1 mM IPTG. After induction the culture was grown for a further 4 hours at 37 ºC with aeration, before being harvested by centrifugation at 4000 g, 4 ºC for 10 minutes. The supernatant was then removed and the pellet resuspended in 30 ml of resuspension buffer (20 mM Tris-HCl [8.0]; 100 mM NaCl; 1 mM EDTA + 1 EDTA-free protease inhibitor tablet [Roche]/50 ml) before being stored at -20 ºC until required for purification.

Expression of Pfu-PCNA was performed using the same protocol as for Pfu-Pol B, except that the plasmid used for transformation of bacterial cells was pET19b-Pfu PCNA (Emptage *et al.*, 2008) and the resuspension buffer (10 mM Tris-HCl [pH 7.4]; 400 mM NaCl; 20 mM imidazole + 1 EDTA-free protease inhibitor tablet [Roche]/50 ml) was altered to enable purification using a HisTrap column (GE Healthcare Lifesciences).

2.5.2: Expression of Pfu-Pol D

WT and mutant versions of pET28a-PfuDP1 (His tagged) and pET22b-PfuDP2 (non tagged) (described in 2.4.12) were used for expression of Pfu-Pol D. The required plasmids were transformed into chemically competent BL21-CodonPlus (DE3)-RIL cells (Stratagene) and selected for on LBagar-$carb_{100}$-kan_{50}-CAM_{34} plates. A single colony was used to inoculate a 50 ml starter culture, which was grown to saturation

overnight under selective conditions before being used to inoculate 2 x 500 ml of LB-$carb_{100}$-kan_{50}-CAM_{34} using 0.5 % (v/v) inoculum. The large-scale culture was grown at 37 ºC with aeration for 1-2 hours (until OD_{600} = 0.6), then induced with 1 mM IPTG. After induction the culture was grown for a further 7 hours at 37 ºC, 150 rpm before being harvested by centrifugation at 4000 g, 4 ºC for 10 minutes. The supernatant was then removed and the pellet resuspended in 30 ml of resuspension buffer (50 mM Tris-HCl [pH8.0], 20 mM imidazole, 500 mM NaCl, 0.1 mM EDTA, 0.5 mM DTT, 10 % glycerol + 1 EDTA-free protease inhibitor tablet [Roche]/50 ml) before being stored at -20 ºC until required for purification.

2.5.3: Purification of Pfu-Pol B

Wild type and mutant variants of Pfu-Pol B were all purified using the same protocol. Frozen/resuspended bacterial pellets were thawed at 37 ºC for roughly 10 minutes then sonicated on ice for 12 minutes in 30 s pulses. DNaseI was added to the cell suspension and incubated at 37 ºC for 30 minutes. After DNaseI digestion, the cell suspension was incubated at 75 ºC for 20 minutes. Insoluble, denatured protein and cellular debris were pelleted by centrifugation at 46000 g, 4 ºC for 80 minutes. The resulting clarified lysate was filtered through a syringe-driven Millex 0.45 µm filter unit (Millipore) to remove any remaining insoluble material. The filtered lysate was applied at 1 ml minute to a 5 ml Heparin column (GE Healthcare Lifesciences), which had previously been washed and equilibrated with buffer A (20 mM Tris-HCl [pH 8.0], 100 mM NaCl, 1 mM EDTA). After extensive washing with buffer A (i.e. until the A280 reading returned to zero) bound protein was eluted at 1 ml/minute using a linear 30 ml gradient of 0-100% buffer B (20 mM Tris-HCl [pH 8.0], 1 M NaCl, 1 mM EDTA). 1 ml fractions were collected and 20 µl of each fraction denatured by heating at 95 ºC for 5 minutes in an equal volume of 2x SDS-PAGE loading buffer (50 mM Tris-HCl [pH 6.8], 2 % SDS, 0.1 % bromophenol blue, 10 % glycerol, 750 mM 2-Mercaptoethanol). Based on the A280 reading of each fraction, 10 µl of denatured sample from each of the peak fractions was analysed by SDS-polyacrylamide gel electrophoresis (PAGE), together with samples of crude lysate, clarified lysate, insoluble fraction, flow-through, wash through and stripped column

fractions. The peak 5 x 1 ml fractions were pooled, then buffer-exchanged into 2x Pol B storage buffer (50 mM Tris-HCl [pH 8.0]; 600 mM NaCl; 2 mM EDTA) by several rounds of centrifugation in an Amicon Ultra 30 kDa cut-off centrifugal filter (Millipore). The concentrated protein suspension was then filter-sterilised using a 0.22 μm syringe-driven filter unit (Millipore), before addition of an equal volume of glycerol and storage at -20 °C.

2.5.4: Purification of Pfu-Pol D

Wild type and mutant variants of Pfu-Pol D were all purified using the same protocol, which was largely based on the protocol used for purification of Pfu-Pol B. All of the initial steps were identical, except that a 5 ml HisTrap column (GE Healthcare Lifesciences) was used in place of the Heparin column previously described. The column was washed and equilibrated with a different buffer A (50 mM Tris-HCl [pH 8.0], 50 mM imidazole, 500 mM NaCl, 0.1 mM EDTA, 0.5 mM DTT, 10% glycerol). The bound protein was also eluted using a different buffer B (50 mM Tris-HCl [pH 8.0], 500 mM imidazole, 500 mM NaCl, 0.1 mM EDTA, 0.5 mM DTT, 10% glycerol). 1 ml fractions were collected and 20 μl of each fraction was denatured and analysed in the same way as for Pfu-Pol B.

SDS-PAGE analysis revealed that an excess of the smaller DP1 subunit was consistently present in the purified heterodimeric enzyme after HisTrap purification. This was presumed to be due to the location of the purification handle on the DP1 subunit and the fact that DP1 appeared to express at a higher level than DP2 during expression tests. A second round of purification was therefore employed, involving gel filtration of the protein through a Superdex200 10/300 GL column (GE healthcare). The protein was first buffer-exchanged into gel filtration buffer (20 mM Tris-HCl [pH 6.5], 400 mM NaCl, 1 mM DTT) and concentrated to a volume of 200 μl. 200 ml of the same, de-gassed buffer was used to wash and equilibrate the column at 0.5 ml/min. The protein was then loaded onto the column and eluted at 0.5 ml/min with three distinct populations of protein visible on the resulting A280 trace. SDS PAGE analysis revealed that the heterodimer eluted in the first population between 20

and 26 minutes after loading. The excess of DP1 subunit appeared to elute between 26 and 32 minutes and a number of degraded products appeared to elute between 32 and 38 minutes.

The re-purified heterodimeric Pfu-Pol D was then buffer-exchanged into 2x Pol D storage buffer (40 mM Tris-HCl [pH 6.5], 400 mM NaCl, 2 mM DTT). Finally, the concentrated protein suspension was filter-sterilised using a 0.22 μm syringe-driven filter unit (Millipore), before addition of an equal volume of glycerol and storage at -20 °C.

2.5.5: Purification of Pfu-PCNA

HisTrap purification of Pfu-PCNA was performed using the same protocol as that described for Pfu-Pol D, except that the wash buffer (10 mM Tris-HCl [pH 7.4], 400 mM NaCl, 50 mM imidazole), elution buffer (10 mM Tris-HCl [pH 7.4], 400 mM NaCl, 500 mM imidazole) and storage buffer (10 mM Tris-HCl [pH 7.4], 200 mM NaCl, 1 mM EDTA) were altered slightly. The purified protein was also stored on ice at 4 °C.

2.5.6: Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Proteins were analysed by SDS-PAGE using separating gels containing 7-15 % Design A Gel 37.5:1 acrylamide:bisacrylamide (National Diagnostics), 375 mM Tris-HCl (pH 8.8), 0.1 % SDS, 0.05 % ammonium persulphate (APS) and 0.05 % N,N,N',N'-tetramethylethylenediamine (TEMED). Stacking gels contained 4 % acrylamide:bisacrylamide (37.5:1), 125 mM Tris-HCl (pH 6.8), 0.1 % SDS, 0.05 % APS and 0.05 % TEMED. Gels were run in 25 mM Tris, 250 mM glycine and 0.5 % SDS with 5 μl of Bio-Rad Precision Plus Protein Dual Color standard loaded in the outermost lane(s). Once fully resolved, gels were stained in 10 % (v/v) acetic acid, 10 % (v/v) isopropanol and 0.25 % (w/v) bromophenol blue on a platform shaker at

room temperature for 30 minutes, then de-stained in 10 % (v/v) acetic acid and 10 % (v/v) isopropanol on a platform shaker at room temperature overnight.

2.5.7: Concentration determination of Pfu-Pol B and Pol D

The concentration of Pfu-Pol B and Pol D was determined using the Beer-Lambert law:

$$C = A_{280}/\varepsilon l$$

where C = protein concentration in mM units; A_{280} = the absorbance of light of wavelength 280 nm. ε = the extinction coefficient of the protein in $mM^{-1}.cm^{-1}$; and l = the path length of the quartz cuvette in cm. Assumed extinction coefficients of proteins used in this study are provided in table 2.4.

Protein	Extinction coefficient at 280 nm ($mM^{-1}cm^{-1}$)
Pfu-Pol B	129.2
Pfu DP1	62.8
Pfu DP2	156.4

Table 2.4: Extinction coefficients of enzymes used in this study at 280 nm in $mM^{-1}.cm^{-1}$. The values were predicted, assuming 50 % of cysteine residues were reduced, using ExPASy ProtParam (http://expasy.org/tools/protparam.html).

2.5.8: Concentration determination of Pfu PCNA

The concentration of Pfu-PCNA was determined by performing Thermoscientific Coomasie Plus (Bradford) Protein Assays, as described by the manufacturer. Absorbance at 280 nm was deemed unsuitable due to the low number of aromatic amino acids in the sequence of Pfu-PCNA, especially tryptophan, and hence the low absorbance of the protein at 280 nm.

2.6: Oligodeoxynucleotide hybridisation, extension, exonuclease and glycosylase assays

2.6.1: Assessment of hybridisation of DNA duplexes

Before undertaking enzymatic activity assays, gel shift assays were performed to establish that oligodeoxynucleotide duplexes were fully annealed under experimental conditions. DNA duplexes were diluted in 1x reaction buffer (as described in chapters 2.6.2 and 2.6.3) supplemented with native gel loading dye (20 mM Tris-HCl [pH 8.0], 8 % glycerol, 0.1 % appropriate dye [orange G, bromophenol blue or xylene cyanol). Mixtures containing 0.1-1 pMol of labelled oligodeoxynucleotide were loaded in each lane and resolved by native gel electrophoresis using acrylamide gels containing 15 - 17 % AccuGel 19:1 acrylamide:bisacrylamide (National Diagnostics), 1x TBE, 0.05 % APS and 0.05 % TEMED. Gels were run in 1x TBE buffer at 4 W per gel, 35 mA, 225 V for 2-4 hours. The resolved products were visualised using a Typhoon™ variable mode imager (GE Healthcare) with default settings for the fluorophore employed and a pixel size of 100 μm.

2.6.2: Pfu-Pol B 3'-5' exonuclease assays

Pfu-Pol B exonuclease assays were performed at 30 ºC using 20 nM primer-template (the sequences of which are provided in table 2.5). Primer-templates were incubated with 100 nM Pfu-PCNA in 20 mM Tris-HCl (pH 8.5), 10 mM KCl, 2 mM $MgSO_4$ and 10 mM $(NH_4)_2SO_4$. The last component to be added was Pfu-Pol B at a final concentration of 100 nM to initiate the reaction. These conditions had previously been shown to be "saturating", i.e. no further increase in activity could be observed by raising the enzyme concentration (Emptage *et al.*, 2008). The reactions were therefore carried out under genuine single turnover conditions. Pfu-Pol B saturates

DNA substrates with single-stranded uracil-containing regions at near-equimolar concentrations. However, at the concentrations listed above, PCNA is required for Pfu-Pol B to fully bind control substrates due to the lower affinity of Pfu-Pol B for substrates that contain only the four canonical bases.

Name	Sequence
Cy5-24	5′-Cy5-GGGGATCCTCTAGAGTCGACCTGC-3′
44t	3′-CCCCTAGGAGATCTCAGCTGGACGTCCGTACGTTCGAACAGAGG-5′
44a	3′-CCCCTAGGAGATCTCAGCTGGACGACCGTTCGTTCGAACAGAGG-5′
44u+1	3′-CCCCTAGGAGATCTCAGCTGGACGUCCGTTCGTTCGAACAGAGG-5′
44u+2	3′-CCCCTAGGAGATCTCAGCTGGACGAUCGTTCGTTCGAACAGAGG-5′
44u+3	3′-CCCCTAGGAGATCTCAGCTGGACGACUGTTCGTTCGAACAGAGG-5′
44u+4	3′-CCCCTAGGAGATCTCAGCTGGACGACCUTTCGTTCGAACAGAGG-5′
44u+9	3′-CCCCTAGGAGATCTCAGCTGGACGACCGTTCGUTCGAACAGAGG-5′
44h+2	3′-CCCCTAGGAGATCTCAGCTGGACGAHCGTTCGTTCGAACAGAGG-5′
44h+4	3′-CCCCTAGGAGATCTCAGCTGGACGACCHTTCGTTCGAACAGAGG-5′
44mismatch	3′-CCCCTAGGAGATCTCAGCTGGACAACCGTTCGTTCGAACAGAGG-5′
44competitor	5′-GGGGATCCTCTAGAGTCGACCTGCTGGCAAGCAAGCTTGTCTCC-3′
54t+20	3′-CCCCTAGGAGATCTCAGCTGGACGACCGTTCGTTCGAACAGAGTACCTGGCTAT-5′
54u+20	3′-CCCCTAGGAGATCTCAGCTGGACGACCGTTCGTTCGAACAGAGUACCTGGCTAT-5′
Cy5-31	5′-Cy5-GGGGATCCTCTAGAGTCGACCTGCAGGGCAA-3′
Cy5-2APA-31	5′-Cy5-GGGGATCCTCTAGAGTCGACCTGCAGGGC2APA-3′
Cy5-A2AP-31	5′-Cy5-GGGGATCCTCTAGAGTCGACCTGCAGGGCA2AP-3′
45t	3′-CCCCTAGGAGATCTCAGCTGGACGTCCCGTTCGTTCGAACAGAGG-5′
45u-1	3′-CCCCTAGGAGATCTCAGCTGGACGTCCCGUTCGTTCGAACAGAGG-5′
45u0	3′-CCCCTAGGAGATCTCAGCTGGACGTCCCGTUCGTTCGAACAGAGG-5′
45u+1	3′-CCCCTAGGAGATCTCAGCTGGACGTCCCGTTUGTTCGAACAGAGG-5′
45u+2	3′-CCCCTAGGAGATCTCAGCTGGACGTCCCGTTCUTTCGAACAGAGG-5′
45u+3	3′-CCCCTAGGAGATCTCAGCTGGACGTCCCGTTCGUTCGAACAGAGG-5′
45u+4 (g)	3'-CCCCTAGGAGATCTCAGCTGGACGTCCCGTTCGTUCGAACAGAGG-5'
45u+4 (t)	3'-CCCCTAGGAGATCTCAGCTGGACGTCCCGTTCTTUCGAACAGAGG-5'
45competitor	5′-GGGGATCCTCTAGAGTCGACCTGCAGGGCAAGCAAGCTTGTCTCC-3′
ss T	5'-Hex-AATAGGTCGATATCGCGAATGG-3'
ss U	5'-Hex-AATAGGTCGATAUCGCGAATGG-3'
ds TT	5'-Hex-AATAGGTCCTATAGGCGAATGG-3' 3'-TTATCCAGGATATCCGCTTACC-5'
ds TU	5'-Hex-AATAGGTCGATAUCGCGAATGG-3' 3′-TTATCCAGCTATAGCGCTTACC-5′
ds UU	5'-Hex-AATAGGTCGATAUCGCGAATGG-3' 3′-TTATCCAGCUATAGCGCTTACC-5′

Table 2.5: A summary of all oligodeoxynucleotides used in this study for primer extension, exonuclease and uracil-DNA glycosylase assays. The name of each oligodeoxynucleotide is listed in column 1 with the sequence in column 2. Mismatched and deaminated bases are highlighted in red.

2.6.3: Pfu-Pol D 3'-5' exonuclease assays

Pfu-Pol D exonuclease and incubation assays were performed at 50 ºC using 20 nM primer-template (the sequences of which are given in table 2.5). Primer-templates were incubated in 10 mM Tris-HCl [pH 9.0], 50 mM KCl, 10 mM $MgCl_2$. The last component to be added was Pfu-Pol D at a final concentration of 140 nM to initiate the reaction. Pfu-Pol D generally binds DNA substrates with higher affinity than Pfu-Pol B, especially primer-templates that do not contain deaminated bases. PCNA was therefore not required in any of these reactions.

2.6.4: Primer extension assays

Primer extension assays were performed with Pfu-Pol B and Pol D using the same conditions as described for exonuclease assays but with 200-400 μM dATP, dCTP, dGTP and dTTP added to the reaction mix.

2.6.5: dGTP incorporation assays

dGTP incorporation assays were performed with Pfu-Pol D (H445A) using the same conditions as for the standard primer-extension assays except that dATP, dCTP and dTTP were excluded from the reaction mix. The reason being that the Cy5-31/45 primer-templates employed (see table 2.5 for sequences) contained template strand cytosine at the +1 position, relative to the primer-template junction. Incorporation rates were determined for dGTP concentrations ranging from 10-400 μM.

2.6.6: dUTP incorporation assays

To test whether or not Pfu-Pol D was capable of incorporating dUTP with the same efficiency as the four canonical bases, primer extension assays were set up using the

same conditions as those described for standard primer extension assays, except that dUTP was substituted in place of dTTP.

2.6.7: Pab-Pol D glycosylase assays

During the course of this PhD collaborative work was performed with Dr Ghislaine Henneke's group based at IFREMER in Brest. The work focused on the family D polymerases of *P. furiosus* and *Pyrococcus abyssi*. The amino acid sequences of the two enzymes' large and small subunits share 86 % and 77 % similarity and in roughly 50 % of cases, the changes are conservative. The activities of the two enzymes were sufficiently similar that the two were used interchangeably. Glycosylase assays were therefore performed using Pab-pol D under identical conditions to those described for exonuclease assays but with both WT and H445A (exo-) versions of Pab-pol D employed.

2.6.8: Termination of activity assays

All extension, exonuclease and glycosylase assays were terminated in the same way. Aliquots of reaction mix were quenched at pre-determined time points by addition of an equal volume of stop buffer. Two stop buffers were employed during the course of this PhD; the first comprised 40 % formamide, 100 mM EDTA, 2 μM competitor oligodeoxynucleotide and 0.1 % dye. The second stop buffer, which was found to improve the resolution of products, comprised 95 % formamide, 10 mM EDTA, 10 mM NaOH, 2 μM competitor oligodeoxynucleotide and 0.1 % dye. The competitor deoxynucleotide that was used was fully complementary to the template strand but did not possess a 5' label. The competitor functioned as a sink for the template DNA, preventing hybridisation of the labelled primer and hence visualisation of double-stranded products. In order to prevent interference with the fluorophore signal, the dye used in each stop buffer was also varied according to the 5' label of the primer. Orange-G dye was employed with Cy5-labelled primers, bromophenol blue and xylene cyanol with Hex-labelled primers and xylene cyanol alone with fluorescein-

labelled primers. After addition of the stop buffer, quenched aliquots were incubated at 95 ºC for 5 minutes, placed on ice for two minutes, spun down, then placed back on ice once more before analysis by denaturing polyacrylamide gel electrophoresis (as described in chapter 2.7.2). This step promoted denaturing of annealed primer-templates and production of single-stranded, labelled primer.

2.6.9: Determination of DNA saturation by polymerases

In order to establish that DNA substrates were fully saturated by the enzymes that were to be tested, labelled oligodeoxynucleotides were incubated with the enzymes under experimental conditions. The reactions were supplemented with native gel loading dye and analysed exactly as described for the hybridisation assays described in chapter 2.6.1, except the acrylamide gels contained only 8 % 19:1 acrylamide:bisacrylamide and 1x TBE.

2.6.10: Analysis of extension and exonuclease products resulting from polymerase action

The denatured, single-stranded, fluorophore-labelled products of extension, exonuclease and incubation assays were resolved by denaturing gel PAGE using identical conditions to the oligodeoxynucleotide annealing tests previously described, except that products were quenched, as described in chapter 2.6.8, prior to loading. Polyacrylamide gels were also supplemented with 8 M urea. Products were typically run at 4-7 W per gel, 35 mA, 225 V for 3-5 hours and then visualised as described in chapter 2.6.1.

2.6.11: Quantification of products of single turnover incorporation and exonuclease assays

After visualisation, the products of exonuclease and dGTP single turnover incorporation assays were quantified using ImageQuantTL software (GE Healthcare). In the dGTP incorporation assays, primer was invariably observed at the +0, +1 and +2 positions, relative to the primer-template junction, despite the template strand containing a guanine base at +2. The intensity of all three bands was therefore quantified and the percentage of extended product, defined as (the intensity of bands at +1 and +2)/(the intensity of all bands), was calculated at each time point. The products of single turnover exonuclease assays were quantified in the same way, except that this time the percentage of full-length primer remaining was measured, as opposed to the percentage of extended product, and was defined as (the intensity of the band at position 0)/(the intensity of all bands).

2.6.12: Analysis of products of single turnover incorporation and exonuclease assays

After quantification, data from the dGTP incorporation assays were plotted to single order exponential curves using Grafit (version 5.0.11) to produce observed single turnover rates (k_{obs}) at a range of concentrations of dGTP. The observed rates were then plotted against dGTP concentration on a secondary plot and fitted to the following equation, again using Grafit version 5.0.11:

$$k_{obs} = \frac{k_{pol}[\text{dATP}]}{K_D + [\text{dATP}]}$$

Where k_{obs} is the observed rate of dGTP incorporation, k_{pol} is the maximum rate of dGTP incorporation (i.e. at saturating dGTP concentration) and K_D is the dissociation constant of the Pfu-Pol D-primer-template complex for dGTP.

In analysing the products of exonuclease assays, data were plotted to single order exponential decay curves, as follows:

$$y = A_0 e^{-kt} + \text{offset}$$

Where y = % full-length primer remaining, A_0 = initial value, k = the rate constant and t = time.

2.7: K_D determination by fluorescence anisotropy

Fluorescence anisotropy measurements and analyses were performed as described for the direct binding titrations by Reid *et al.*, 2001 unless otherwise stated. All Hex-labelled oligodeoxynucleotides were annealed at 1:1.3 molar ratios (confirmed by native gel mobility shift assays) and employed at a final concentration of 1 nM. In the case of Pfu-Pol B binding assays, oligodeoxynucleotides were suspended in one of three buffers. The first contained 10 mM Tris-HCl (pH 7.5), 100 mM NaCl, 1 mM EDTA and 0.1 mg/ml of bovine serum albumin (BSA), as described by Gill *et al.*, 2007. Alternatively, oligodeoxynucleotides were suspended in 20 mM Tris-HCl (pH 8.5), 10mM KCl, 10 mM $(NH_4)_2SO_4$ and either 2 mM $MgCl_2$ or 2 mM $CaCl_2$ (NB/ the $[NH_4]_2SO_4$ was omitted when $CaCl_2$ was used). Pfu-Pol B D215A (exo-) was then titrated directly into the solution at concentrations ranging from 1 to 26 nM. In the case of Pfu-Pol D binding assays, oligodeoxynucleotides were suspended in 20 mM Tris-HCl (pH 8.8), 10 mM KCl, 1 mM DTT and 0.1 mg/ml of bovine serum albumin (BSA). Pfu-Pol D H445A (exo-) was then titrated directly into solution at concentrations ranging from 1 to 50 nM.

All oligodeoxynucleotides used in fluorescence anisotropy binding experiments during the course of this PhD are listed in table 2.6.

Name	Sequence
^{2A}PA-31	5'-GGGGATCCTCTAGAGTCGACCTGCAGGGCPA-3'
A^{2A}P-31	5'-GGGGATCCTCTAGAGTCGACCTGCAGGGCAP-3'
45t	3'-CCCCTAGGAGATCTCAGCTGGACGTCCCGTTCTTTCGAACAGAGG-5'
45u+2	3'-CCCCTAGGAGATCTCAGCTGGACGTCCCGTTCUTTCGAACAGAGG-5'
45u+4 (g)	3'-CCCCTAGGAGATCTCAGCTGGACGTCCCGTTCGTUCGAACAGAGG-5'
45u+4	3'-CCCCTAGGAGATCTCAGCTGGACGTCCCGTTCTTUCGAACAGAGG-5'
Hex 40t	5'-Hex-TTTCTGGTTCCAGCTGGACCATTCGCCTATAGGACCTATT-3'
Hex 40u	5'-Hex-TTTCTGGTUCCAGCTGGACCATTCGCCTATAGGACCTATT-3'
+10	3'-GGTAAGCGGATATCCTGGATAA-5'
+4	3'-CGACCTGGTAAGCGGATATCCTGGATAA-5'
+2	3'-GTCGACCTGGTAAGCGGATATCCTGGATAA-5'
+1	3'-GGTCGACCTGGTAAGCGGATATCCTGGATAA-5'
-1	3'-AAGGTCGACCTGGTAAGCGGATATCCTGGATAA-5'
Comp t	3'-AAAGACCAAGGTCGACCTGGTAAGCGGATATCCTGGATAA-5'
Comp u	3'-AAAGACCAAGGTCGACCTGGTAAGCGGAUATCCTGGATAA-5'

Table 2.6: A summary of all oligodeoxynucleotides used in this thesis for fluorescence studies. Uracil residues are indicated by a red U. 2-aminopurine (2AP) residues are indicated by a green P.

2.8: Steady state fluorescence measurements of 2-aminopurine-containing DNA primer-templates and DNA-enzyme complexes

Fluorescence emission spectra of 2-aminopurine (2AP)-containing DNA primer-templates and DNA-enzyme complexes were measured using a Quartz cuvette (Hellma) with a 100 µl volume at 30 °C. Primer-templates were used containing the 2APA-31 and A2AP-31 primers annealed to the 45u+2, 45u+4 and 45u+4 #2 templates (all described in table 2.6) at a 1:2 molar ratio with the template strand always in excess. Assays were performed in a total volume of 100 µl containing 2 µM primer-template suspended in one of two buffers containing 20 mM Tris-HCl (pH 8.5), 10mM KCl, 10 mM $(NH_4)_2SO_4$ and either 2 mM $MgCl_2$ or 2 mM $CaCl_2$ (NB/ the $[NH_4]_2SO_4$ was omitted when $CaCl_2$ was used). Where DNA-enzyme complexes were measured Pfu-Pol B D215A was added to a final volume of 8 µM. Samples were excited at 315 nm using a Cary Eclipse Fluorescence Spectrophotometer (Varian) connected to an integrated Cary temperature controller (Varian). Emission

spectra were measured between 330 nm and 400 nm and the data collected using ADL Shell software (Agilent Technologies, Santa Clara, USA). Blank spectra of the buffers on their own were recorded to enable measurement of the raman band, which was subsequently subtracted from the primer-template spectra. Similarly, blank spectra of Pfu-Pol B D215A suspended in each buffer were recorded and subtracted from measurements of the primer-template-enzyme complexes.

2.9: Time-resolved fluorescence measurements of 2AP-containing DNA primer-templates and DNA-enzyme complexes

Time-resolved fluorescence measurements of 2AP-containing DNA primer-templates and DNA-enzyme complexes were performed using the same buffers, temperature, DNA and enzyme concentrations as those described for the steady state measurements in chapter 2.8. Fluorescence decays were acquired by time-correlated single photon counting using a Ti-Sapphire femtosecond laser system as the excitation source (Neely *et al.*, 2005; Youngblood *et al.*, 2008). The excitation wavelength was 315 nm with pulses of ≈200 fs at 4.75 MHz repletion rate. The instrument response was ≈85 ps FWHM and both monochromator slit widths were 7.77 nm, giving a spectral resolution of 14 nm. Decays were collected at three emission wavelengths (365, 380 and 395 nm) and fitted to the function:

$$I(t) = B + \sum_i A_i e^{\frac{-t}{\tau_i}}$$

where *I(t)* is the intensity at time t, *B* is the background level (i.e. the dark count of the detector), A_i is the fractional amplitude of the i^{th} lifetime component and τ_i is the fluorescence lifetime of the i^{th} species (i.e. the time taken for intensity to fall to 1/e of its initial magnitude).

The decays for each sample were analysed globally; the set of decays for one sample at different emission wavelengths was fitted simultaneously with lifetimes, $\tau_{i,}$ as common parameters. The quality of fits was assessed by the χ^2 parameter, visual examination of the fitted function relative to the data and from a plot of the residuals. Typically, a χ^2 value between 1.0 and 1.2 indicates an acceptable fit. All of the decay curves required four exponential components to give a satisfactory fit, as observed in previous studies (Neely *et al.*, 2005; Youngblood *et al.*, 2008).

2.10: Growth media and culturing conditions of *Thermococcus kodakarensis*

2.10.1: Stock solutions for *T. kod*

200x stock of *T. kod*-specific trace minerals (per litre):

0.5 g $MnSO_4.H_2O$
0.1 g $CoCl_2.H_2O$
0.1 g $ZnSO_4.7H_2O$
10 mg $CuSO_4.5\ H_2O$
10 mg $AlK(SO_4)_2.12H_2O$
10 mg H_3BO_4
10 mg $Na_2MoO_4.2H_2O$

Dissolved in double distilled H_2O, adjusted to 1 l and stored at 4˚C.

2x artificial seawater (ASW) (per litre):
40 g NaCl
6 g $MgCl_2.6H_2O$
12 g $MgSO_4.7H_2O$
2 g $(NH_4)_2SO_4$

0.4 g $NaHCO_3$

0.6 g $CaCl_2.2H_2O$

1.0 g KCl

0.84 g KH_2PO_4

100 mg NaBr

40 mg $SrCl_2$-$6H_2O$

20 mg Fe $(NH_4)_2(SO_4)_2.6H_2O$

Dissolved in double distilled H_2O, adjusted to 1 l and stored at 4˚C.

200x Amino Acid Mix (per 200 ml):

1 g cys, glu, gly

500 mg arg, pro

400 mg asn, his, ile, leu, lys, thr, tyr

300 mg ala, met, phe, ser, trp

200 mg val, asp, gln

Dissolved in double distilled H_2O, adjusted to 200 ml and stored at room temperature.

This is in fact a complete list and, as such, tryptophan was omitted from all of the plates that were used. The reason for this is that all of the *T. kodakarensis* strains that were used were based on KW128, which is a tryptophan auxotroph, enabling us to use the TrpE gene as a selective marker for transformants that had undergone allelic exchange.

1000x *T. kod*-specific vitamin stock mixture (per 50 ml):

10 mg niacin

4 mg biotin

10 mg pantothenate

10 mg lipoic acid

4 mg folic acid

10 mg p-aminobenzoic acid

10 mg thiamine

10 mg riboflavin

10 mg pyridoxine

10 mg cobalamin

Dissolved in double distilled H_2O, adjusted to 50 ml and stored at room temperature.

Polysulfides solution

10 g Na_2S-$9H_2O$

3 g S^0

Dissolved in double distilled H_2O, adjusted to 15 ml and stored at room temperature.

2.10.2: Plating of *T. kod*

1 g of Gelzan/Gel-rite (Sigma) was dissolved in 50 ml of 18 MΩ H_2O and sealed in a serum bottle. In a separate serum bottle 500 µl of 200x *T. kod*-specific trace mineral stock was added to 50 ml of 2x ASW and sealed inside. Both serum bottles were then autoclaved at 120 C for at least 20 minutes. The bottles containing both components were removed from the autoclave promptly and brought to the anaerobic chamber. 0.5 ml of 200x amino acid stock (lacking tryptophan) was added, followed by 100 µl of the vitamin stock and 200 µl of polysulfides solution. All 50 ml of the gel-rite solution was then added and 4 x 25 ml plates were poured into thick glass petri dishes. Where mutants for TK0664 were being screened, 6-methylpurine (6MP) was added to the plate media at the same time as the vitamin stock at a final concentration of 100 µM.

2.10.3: Liquid culturing of *T. kod*

ASW•YT (yeast tryptone) medium (per litre)

20.0 g NaCl

6.0 g $MgSO_4.7H_2O$

3.0 g $MgCl_2.6H_2O$

5.0 g Yeast Extract

5.0 g Tryptone

1.0 g $NH_4(SO_4)_2$

0.2 g $NaHCO_3$

0.3 g $CaCl_2.2H_2O$

0.5 g KCl

0.42 g KH_2PO_4

50 mg NaBr

20 mg $SrCl_2.6H_2O$

10 mg $Fe(NH_4)_2(SO_4)_2$ $.6H_2O$

5.0 ml 200x *T. kod*-specific Trace Mineral

1 l boiled H_2O

100 ml aliquots of medium were dispensed into serum bottles in an anaerobic atmosphere. Bottles were then sealed and autoclaved at for at least 20 minutes. Either a swab from a single colony, or a 1/100 inoculum of starter culture, was used to set up liquid cultures of *T. kod*. Before inoculation *T. kod*-specific vitamins were added, as well as solid sulphur to ≈ 0.25 % w/v.

2.11: Transformation of *T. kod*

For production of mutant strains TR1, TR2, TR3 and TR4, the four plasmids described in chapter 2.4.11 were transformed into a KW128 strain of *T. kod*. KW128 is a tryptophan auxotrophic mutant with a disrupted TrpE gene. The TrpE gene contained within each of the pTR-pTR4 plasmids allowed selection for desired transformants.

10 ml cultures of *T. kod* KW128 were grown in ASW-YT medium to mid-exponential phase for roughly 10-12 hours at 85˚C then harvested, under anaerobic conditions, by centrifugation at 17000 g for 5 minutes at 4 ˚C. The supernatant was carefully

discarded and the cell pellet resuspended in 200 µl of 0.8 % ASW. The cell suspension was kept on ice for 30 minutes before incubation with 1-3 µg of donor DNA for an additional 60 minutes. The cells were heat-shocked at 85 ºC for 45 s, then placed back on ice for a further 10 minutes. The cells were transferred to 3 ml of ASW-YT media supplemented with vitamins and sulphur and cultivated for 2 hours at 85 ºC. The cultures were then harvested by centrifugation at 10,000 g for 2 minutes. The supernatant was removed from each tube, the pellet resuspended in 200 µl of 0.8x ASW each and plated on Trp- plates. Plates were then incubated at 85 ºC for 3-5 days or until colonies appeared. Typical transformation efficiency was approximately 1-2000 colonies/g of DNA/$1.5x10^8$ cells.

After 3-5 days individual colonies were resuspended in 5 µl of 0.8x ASW, which was then used to perform serial dilutions onto a fresh Trp- plate. This re-spotting step helped to reduce the likelihood of KW128 contamination. After another 2-4 days, single re-spotted colonies were used to set up 5 ml cultures of ASW-YT medium, which were grown to saturation at 85 ºC for roughly 15 hours. 1.5 ml of each culture was removed and subjected to genomic DNA extraction, as described in chapter 2.12. Genomic DNA from each of the amplified transformants was then used to perform diagnostic PCRs using the TK0001 2100RP, TK0001 seq, TK2303 1300FP, TrpE FP and TrpE RP primers (described in table 2.3) to amplify fragments beginning within the TrpE marker and extending beyond the cloned region of either TK0001 or TK2303. After agarose gel electrophoretic analysis, the presence of both products was used to confirm whether or not cultures were derived from single positive transformants that contained the TrpE marker. Sequencing of PCR products was performed to confirm the presence of the desired mutations in TK0001, as well as the absence of any extraneous mutations. Finally the entire region, containing the TrpE marker and the two cloned pieces of plasmid DNA was amplified to confirm the absence of wild-type KW128 DNA.

2.12: *T. kod* genomic DNA preparation

T. kod cultures were grown to saturation at 85 ºC for 15 hours. 1.5 ml of saturated culture was then harvested by centrifugation at 10,000 g for 5 minutes, the supernatant was removed and the pellet resuspended in 200 µl of resuspension buffer (20 mM Tris-HCl [pH 8.0], 5 mM EDTA, 10 % sucrose). The cell suspension was then placed briefly in a bench top vortexer. 50 µl of 10% SDS was added to each cell suspension before returning to the vortexer. At this point 3.5 µl of proteinase K (20 mg/ml) was added and resuspended. The cell suspension was incubated at 55 ºC for 30 minutes. After 30 minutes NaCl was added to a final concentration of 1 M, the whole mixture was then vortexed and placed on ice for 5 minutes. Precipitated protein and cellular debris were pelleted by centrifugation at 14000 g, 4 ºC for 10 minutes. The supernatant was then transferred to a fresh Eppendorf tube containing 320 µl of isopropanol and mixed. Precipitated DNA was pelleted by centrifugation at 14000 g for 30 minutes. The supernatant was then removed and any residual liquid allowed to evaporate in a heat block. Precipitated DNA was resuspended in 20 µl of 10 mM Tris-HCl (pH 8.0) and stored at -20 ºC.

2.13: Selection of TK0664 mutants

In order to produce mutagenic profiles for each of our mutant *T. kod* strains, a screen based on the TK0664 gene was employed. Firstly, to ensure that the starter cultures were monoclonal and originated from cells that were wild-type for TK0664, serial dilutions of each strain were plated out and grown on Trp- plates. Once visible, single colonies were resuspended in 6 µl of 0.8x ASW with 3 µl of the resulting suspension plated on Trp-/6Mp plates and the remaining 3 µl plated, in serial dilutions, on Trp- plates. After 3-5 days growth at 85 ºC, if colonies were present on the Trp- plates, but not on the corresponding Trp-/6Mp plates, single colonies were picked from the Trp- plates and used to set up 2 ml cultures of ASW-YT medium supplemented with *T. kod*-specific vitamins and solid sulphur. Cultures were grown at 85 ºC for roughly 15 hours, then harvested by centrifugation at 10000 g for 5 minutes. The cell pellet was resuspended in 50 µl of 0.8x ASW and plated onto Trp-/6Mp plates. After 3-5 days

single colonies, which contained knockout mutations in TK0664, were used to inoculate 1.5 ml cultures of ASW-YT medium. The cultures were then grown to saturation and subjected to genomic DNA extraction, as described in chapter 2.12. PCRs were performed on the purified DNA using the TK0664 FP and RP primers (described in table 2.3) to produce an 839 bp product covering the TK0664 gene, as well as the promoter region and a further 100 bp downstream from the stop codon. The same primers were used to perform sequencing reactions covering the full length of the TK0664 open reading frame (ORF) and surrounding regions. Sequencing results were then used to create profiles of the types of mutations produced by each strain.

Chapter 3

Influence of deaminated bases on the 3'-5' exonucleolytic activity of archaeal family B DNA polymerases

3.1: Background

As discussed in the introduction, the structures and mechanisms responsible for selective binding of deaminated bases by archaeal family B polymerases have been well characterised. However, at present the events in the deaminated base recognition pathway that occur post-stalling remain largely unknown. It is well known that most replicative polymerases are capable of adopting two distinct conformations; one associated with polymerisation and the other with proofreading exonuclease activity (Freemont *et al.*, 1988; Baker and Bell, 1998). Upon commencing the work described in this chapter, only one crystal structure of an archaeal family B polymerase in complex with a uracil-containing DNA substrate was available (Firbank *et al.*, 2008). The structure, shown in figure 3.1, is that of the family B DNA polymerase of *Thermococcus gorgonarius* bound to a primer-template mimic with uracil located in the single-stranded template region 4 bases downstream from the primer-template junction (i.e. at the preferred +4 position). Interestingly, superimposition of the *T. gorgonarius* Pol B structure with previously solved co-crystal structures of an α-type DNA polymerase and the bacteriophage RB69 polymerase appears to suggest that Tgo Pol B adopts an editing, not a polymerising, conformation when bound to the U+4 substrate, as shown in figure 3.1. Unlike in most previous editing conformations, however, the 3' terminus of the primer does not protrude into the 3'-5' exonuclease active site. This observation was highly unexpected since the duplexed region of the DNA hairpin contains only *bona fide* Watson-Crick base pairs. Prior to the production of this structure, no link between deaminated base stalling and 3'-5' exonucleolytic activity had been suspected. In fact, the majority of studies on deaminated base recognition by archaeal family B polymerases have tended to focus on enzymes deficient for 3'-5' exonuclease activity to prevent unwanted degradation of DNA substrates (Greagg *et al.*, 1999; Fogg *et al.*, 2002; Shuttleworth *et al.*, 2004). In order to probe the effect of template strand deaminated bases on the proofreading exonuclease activity of archaeal Pol B, a series of assays was performed, the results of which are fully described in chapters 3.2-3.6.

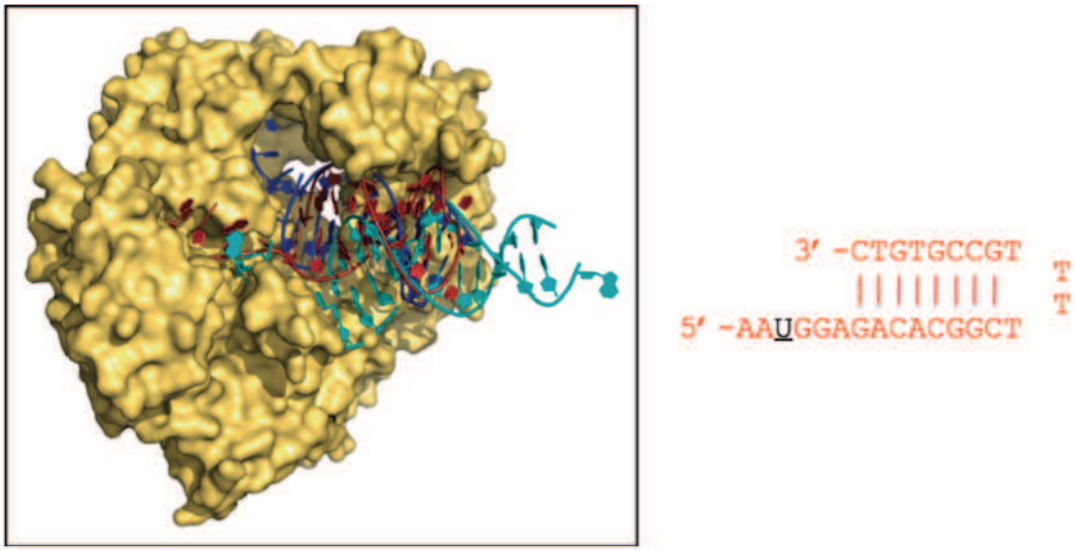


Figure 3.1: Crystal structure of the family B DNA polymerase of *Thermococcus gorgonarius* (yellow) in complex with a U+4 primer-template mimic (red) (Firbank *et al.*, 2008). The predicted positions of DNA bound in the polymerising (cyan) and editing (blue) modes have been superimposed based on co-crystal structures of an α-type DNA polymerase and the bacteriophage RB69 polymerase (Franklin *et al.*, 2001; Shamoo and Steitz, 1999). The sequence of the primer-template mimic is shown on the right with the uracil residue represented by an underlined U.

3.2: 3'-5' exonucleolytic activity of archaeal family B DNA polymerases upon encountering template strand deaminated bases using primer-templates that possess gc clamps

Based on the structural evidence presented in chapter 3.1, it was hypothesised that a stimulation of 3'-5' exonucleolytic activity might occur as archaeal family B polymerases approach template strand deaminated bases. In order to test whether or not this is the case, exonuclease assays were performed using wild type Pfu-Pol B. A series of Cy5-labelled 24/44mer primer-templates was designed, synthesised and annealed, the sequences of which are provided in table 3.1. The substrates were designed to contain uracil and hypoxanthine at a variety of positions relative to the

primer-template junction, ranging from +1 to +4. Mismatched and fully complementary primer-templates containing only the four canonical bases were also tested for reference. All of the tested substrates contained a guanine residue followed by a cytosine at the 3' terminus of the primer strand. In all cases, except for the mismatched substrates, these residues formed base pairs with cytosine and guanine residues located within the template strand. The sequences of all of the primer-templates are provided in table 3.1.

Designation	**Sequence**
Control (gc/cg)	5′-Cy5-GGGGATCCTCTAGAGTCGACCTGC-3′ 3′-CCCCTAGGAGATCTCAGCTGGACGACCGTTCGTTCGAACAGAGG-5′
Mismatch (gc/cg)	5′-Cy5-GGGGATCCTCTAGAGTCGACCTGC-3′ 3′-CCCCTAGGAGATCTCAGCTGGACAACCGTTCGTTCGAACAGAGG-5′
U+1 (gc/cg)	5′-Cy5-GGGGATCCTCTAGAGTCGACCTGC-3′ 3′-CCCCTAGGAGATCTCAGCTGGACGUCCGTTCGTTCGAACAGAGG-5′
U+2 (gc/cg)	5′-Cy5-GGGGATCCTCTAGAGTCGACCTGC-3′ 3′-CCCCTAGGAGATCTCAGCTGGACGAUCGTTCGTTCGAACAGAGG-5′
U+3 (gc/cg)	5′-Cy5-GGGGATCCTCTAGAGTCGACCTGC-3′ 3′-CCCCTAGGAGATCTCAGCTGGACGACUGTTCGTTCGAACAGAGG-5′
U+4 (gc/cg)	5′-Cy5-GGGGATCCTCTAGAGTCGACCTGC-3′ 3′-CCCCTAGGAGATCTCAGCTGGACGACCUTTCGTTCGAACAGAGG-5′
I+2 (gc/cg)	5′-Cy5-GGGGATCCTCTAGAGTCGACCTGC-3′ 3′-CCCCTAGGAGATCTCAGCTGGACGACIGTTCGTTCGAACAGAGG-5′
I+4 (gc/cg)	5′-Cy5-GGGGATCCTCTAGAGTCGACCTGC-3′ 3′-CCCCTAGGAGATCTCAGCTGGACGACCGITCGTTCGAACAGAGG-5′

Table 3.1: Names and sequences of primer-templates used in chapters 3.2 and 3.4 of this thesis. Uracil (U), hypoxanthine (I) and mismatched bases (A) are highlighted in red.

It has previously been shown that Pfu-Pol B tightly binds single-stranded uracil-containing substrates (Shuttleworth *et al.*, 2004; Gill *et al.*, 2007). As such, Pfu-Pol B is capable of saturating such substrates at near equimolar concentrations. In order to promote full binding of DNA substrates that contain only the four canonical bases, however, the processivity factor PCNA must be added to reactions to increase the enzyme's affinity. The exonuclease and extension assays presented in chapter 3 were therefore carried out in the presence of PCNA using conditions that have previously been shown to result in full binding (Emptage *et al.*, 2008). The reactions were therefore performed under genuine single turnover conditions. Gel images resulting from 3'-5' exonuclease assays using the substrates shown in table 3.1 can be seen in figure 3.2.

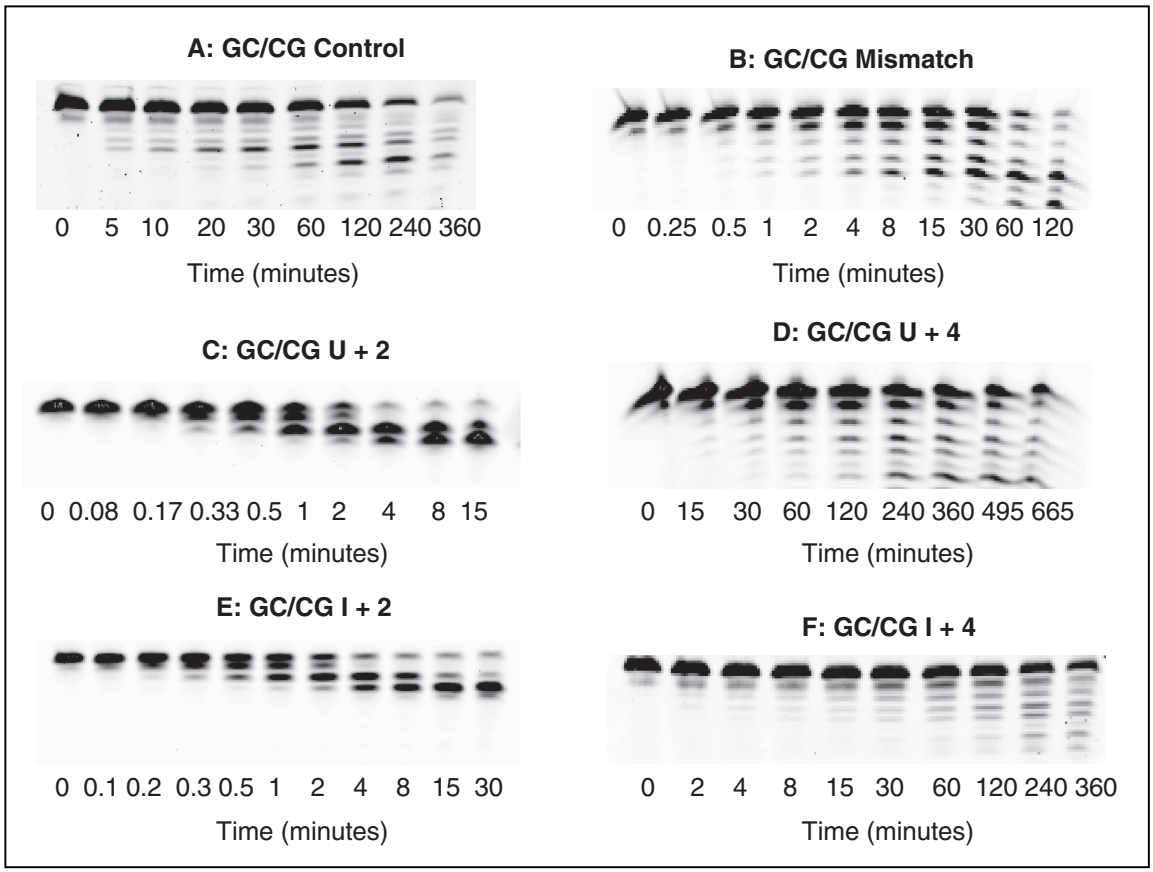


Figure 3.2: Gel images of exonuclease assays performed using Pfu-Pol B with six different 24/44mer DNA primer-templates, the names of which are given above each image. The sequences of all of the primer-templates are provided in table 3.1. The marker, which appears in the lane furthest to the left for each assay, is the non-degraded Cy5-labelled primer itself.

Reference to figure 3.2 reveals that the control primer-template is degraded by Pfu-Pol B with a half-life of approximately 2 hours, which was significantly slower than the observed rate of degradation for the mismatched primer-template. With the mismatched primer-template the majority of the full-length primer had been degraded after 60 minutes. Remarkably, with the U+2 and I+2 primer-templates most of the full-length primer was degraded within 4 minutes. In contrast, however, the slowest rate of degradation was observed when uracil or hypoxanthine was situated at the +4 position, with much of the full-length primer persisting after 6 hours.

By quantifying the intensity of each of the bands in gel images such as those shown figure 3.2, it was possible to calculate the total percentage of full-length primer remaining at each time point. The resulting data was then plotted and fitted to single order exponential decay curves, a selection of which can be seen in figure 3.3.

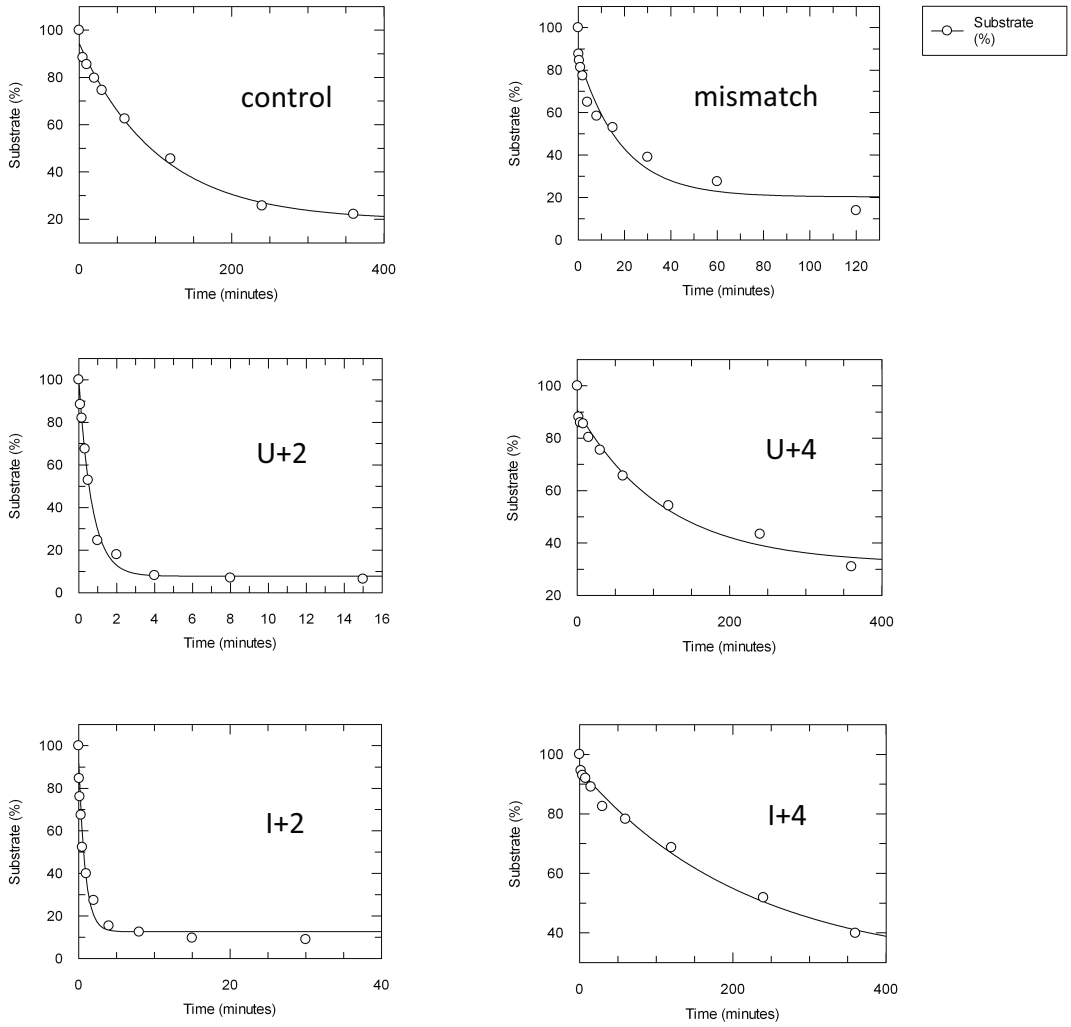


Figure 3.3: First order reaction curves of the 3'-5' exonucleolytic activity of Pfu-Pol B with the primer-templates described in table 3.1. The curves were plotted using quantified data from gel images, such as those shown in figure 3.2. The data were subsequently fitted to single order exponential decay curves.

Using the plots shown above, it was possible to calculate the single turnover rate constants (k_{exo}s) for the 3'-5' exonucleolytic activity of Pfu-Pol B with each of the six

substrates (as described in chapter 2.6.12). The rate constants observed with each primer-template are summarised in table 3.2. The relative rate, defined as the k_{exo} for a given primer-template divided by the k_{exo} for the relevant control, is also provided for each substrate.

Primer-template	Repeat	Rate Constant (min^{-1})	Mean Rate Constant (min^{-1})	Standard Deviation	Relative rate
Control (gc/cg)	1 2 3	0.0096 0.012 0.011	0.011	0.001	1
Mismatch (gc/cg)	1 2 3	0.054 0.060 0.068	0.061	0.007	5.5
U+2 (gc/cg)	1 2 3	1.4 2.0 1.9	1.8	0.3	163
U+4 (gc/cg)	1 2 3	0.0074 0.0088 0.0079	0.0080	0.0007	0.7
I+2 (gc/cg)	1 2 3	1.2 1.8 1.6	1.5	0.3	136
I+4 (gc/cg)	1 2 3	0.0045 0.013 0.0088	0.0086	0.004	0.8

Table 3.2: A summary of the single turnover rate constants calculated for the 3'-5' exonucleolytic activity of Pfu-Pol B with each of six tested substrates. The sequences of the substrates are all provided in table 3.3.

Reference to table 3.2 reveals a substantial 163-fold increase in 3'-5' exonucleolytic activity with the U+2 primer-template, compared to the control. Interestingly, this increase in activity is considerably greater than the 5.5-fold increase that was observed the mismatched primer-template. The U+4 and I+4 primer-templates were the only substrates to show a decrease in 3'-5' exonucleolytic activity, compared to the control. The mean k_{cat}s for the U+4 and I+4 primer-templates were 30 % and 20 % lower, respectively, than that of the control.

3.3: 3'-5' exonucleolytic activity of archaeal family B DNA polymerases upon encountering template strand deaminated bases using primer-templates that do not possess gc clamps

All of the data presented in chapter 3.3 were produced by Mr Henry J Russell, a former undergraduate project student in the Connolly lab.

In order to test whether or not the bases at the 3' terminus of the primer strand had any influence on the stimulation in exonucleolytic activity previously observed, exonuclease assays were repeated using a different set of primer-templates, all of which contained two adenine residues at the 3' terminus of the primer strand. In all cases, except the mismatched primer-template, these terminal residues formed base pairs with two thymine residues located in the template strand. The sequences of all of the primer-templates are shown in table 3.3

Designation	Sequence
Control (aa/tt)	5'-Cy5-GGGGATCCTCTAGAGTCGACCTGCAGGGCAA-3' 3'-CCCCTAGGAGATCTCAGCTGGACGTCCCGTTCGTTCGAACAGAGG-5'
Mismatch (aa/tt)	5'-Cy5-GGGGATCCTCTAGAGTCGACCTGCAGGGCAA-3' 3'-CCCCTAGGAGATCTCAGCTGGACGTCCCGTCCGTTCGAACAGAGG-5'
U-1 (aa/tt)	5'-Cy5-GGGGATCCTCTAGAGTCGACCTGCAGGGCAA-3' 3'-CCCCTAGGAGATCTCAGCTGGACGTCCCGUTCGTTCGAACAGAGG-5'
U0 (aa/tt)	5'-Cy5-GGGGATCCTCTAGAGTCGACCTGCAGGGCAA-3' 3'-CCCCTAGGAGATCTCAGCTGGACGTCCCGTUCGTTCGAACAGAGG-5'
U+1 (aa/tt)	5'-Cy5-GGGGATCCTCTAGAGTCGACCTGCAGGGCAA-3' 3'-CCCCTAGGAGATCTCAGCTGGACGTCCCGTTUGTTCGAACAGAGG-5'
U+2 (aa/tt)	5'-Cy5-GGGGATCCTCTAGAGTCGACCTGCAGGGCAA-3' 3'-CCCCTAGGAGATCTCAGCTGGACGTCCCGTTCUTTCGAACAGAGG-5'
U+3 (aa/tt)	5'-Cy5-GGGGATCCTCTAGAGTCGACCTGCAGGGCAA-3' 3'-CCCCTAGGAGATCTCAGCTGGACGTCCCGTTCGUTCGAACAGAGG-5'
U+4 (aa/tt)	5'-Cy5-GGGGATCCTCTAGAGTCGACCTGCAGGGCAA-3' 3'-CCCCTAGGAGATCTCAGCTGGACGTCCCGTTCGTUCGAACAGAGG-5'

Table 3.3: Names and sequences of primer-templates used in chapters 3.3 and 3.5 of this thesis. Uracil and mismatched bases are highlighted in red.

This alternative set of primer-templates were subjected to exonuclease assays using exactly the same conditions as those described in chapter 3.2. The results were subsequently analysed in the same manner to produce the rate constants shown in table 3.4. Based on the results of preliminary experiments, it was apparent that the rates of degradation for the U0, U+1, U+2 and U+3 primer-templates were all very similar, likewise for the U-1 and control primer-templates. It was therefore decided that quantitative analyses would be performed only on the control, mismatched, U+2 and U+4 primer-templates.

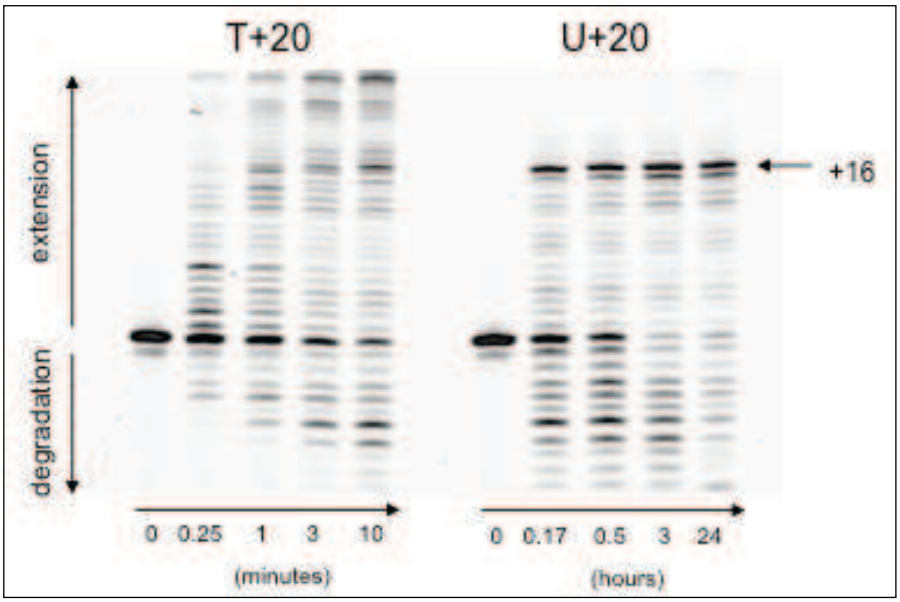
Primer-template	Mean Rate Constant (min^{-1})	Standard Deviation	Relative rate
Control (aa/tt)	0.84	0.06	1
Mismatch (aa/tt)	9	2	10.7
U+2 (aa/tt)	48	12	57
U+4 (aa/tt)	0.42	0.06	0.5

Table 3.4: A summary of the single turnover rate constants calculated for the 3'-5' exonucleolytic activity of Pfu-Pol B with each of four tested substrates. The sequences of the substrates are all provided in table 3.3.

3.4: Replicative bypass of template strand uracil and hypoxanthine by wild type and 3'-5' exonuclease-deficient strains of Pfu-Pol B

Read-ahead recognition and replicative stalling by archaeal family B polymerases can be simply demonstrated by performing primer extension assays, as shown in figures 3.4 and 3.5. In both sets of assays two Cy5-labelled primer-templates have been

employed, the sequences of which are provided within figure 3.4. Critically, the primer-templates differ by a single base at the +20 position in the template strand, relative to the primer-template junction, with one containing a thymine residue (T+20) and the other a uracil (U+20).




```
5'-Cy5-GGGGATCCTCTAGAGTCGACCTGC-3'
    3'-CCCCTAGGAGATCTCAGCTGGACGACCGTTCGTTCGAACAGAG(T/U)ACCTGGCTAT-5'
```

Figure 3.4: PAGE analysis of the products of primer extension assays performed using wild type Pfu-Pol B with two different primer-templates, "T+20" and "U+20", the sequences of which are shown in the box.

Figure 3.4 reveals that after 10 minutes the T+20 reaction is more or less complete, with the majority of the Cy5-labelled primer fully-extended. However, with the U+20 primer-template a very prominent stall can be seen at the +16 position, i.e. four bases upstream from the uracil residue. Even after 24 hours virtually no fully extended product is seen in the U+20 assay. Clearly one can deduce from these results that deaminated base recognition and subsequent stalling represents an extremely efficient means of preventing replicative bypass of uracil by archaeal family B polymerases.

In order to probe the role of Pfu-Pol B's 3'-5' exonucleolytic activity in preventing replicative bypass of deaminated bases, the primer extension assays described in

figure 3.4 were repeated using a D215A mutant version of Pfu-Pol B. This mutation has previously been shown to almost completely eliminate 3'-5' exonucleolytic activity (Evans *et al.*, 2000). The resulting gel images can be seen in figure 3.5.

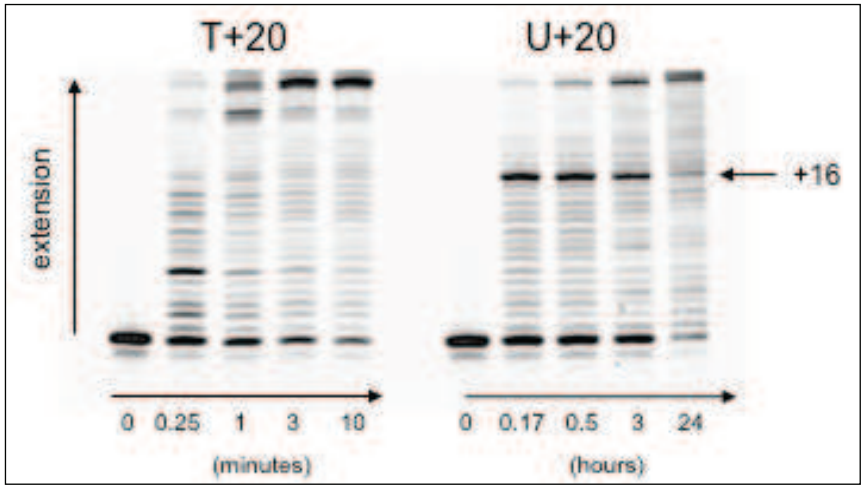


Figure 3.5: PAGE analysis of the products of two primer extension assays performed using a 3'-5' exonuclease-deficient D215A mutant version of Pfu-Pol B with two different primer-templates, the sequences of which are provided in figure 3.4.

Comparison of the wild type and D215A T+20 extension reactions (figures 3.4 and 3.5) reveals that the two enzymes generate very similar profiles of products, except for the presence or absence of degradation products at the bottom of the images. Comparison of the U+20 reactions, however, revealed that, in stark contrast to the wild type enzyme, the D215A mutant produces fully extended product after just 10 minutes, with the majority of the primer fully extended by the final time point.

From figures 3.4 and 3.5 one can clearly conclude that the 3'-5' exonucleolytic activity of archaeal Pol B plays a significant role in maintaining the enzyme at the preferred -4 position upon encountering template strand uracil. This finding adds further weight to the observation that 3'-5' exonucleolytic activity plays a major role in the deaminated base recognition pathway.

In order to further confirm the role of 3'-5' exonucleolytic activity in the deaminated base recognition pathway, primer extension assays were performed using the primer-

templates previously employed in chapters 3.2 and 3.3. The results of these assays are presented in figure 3.6.

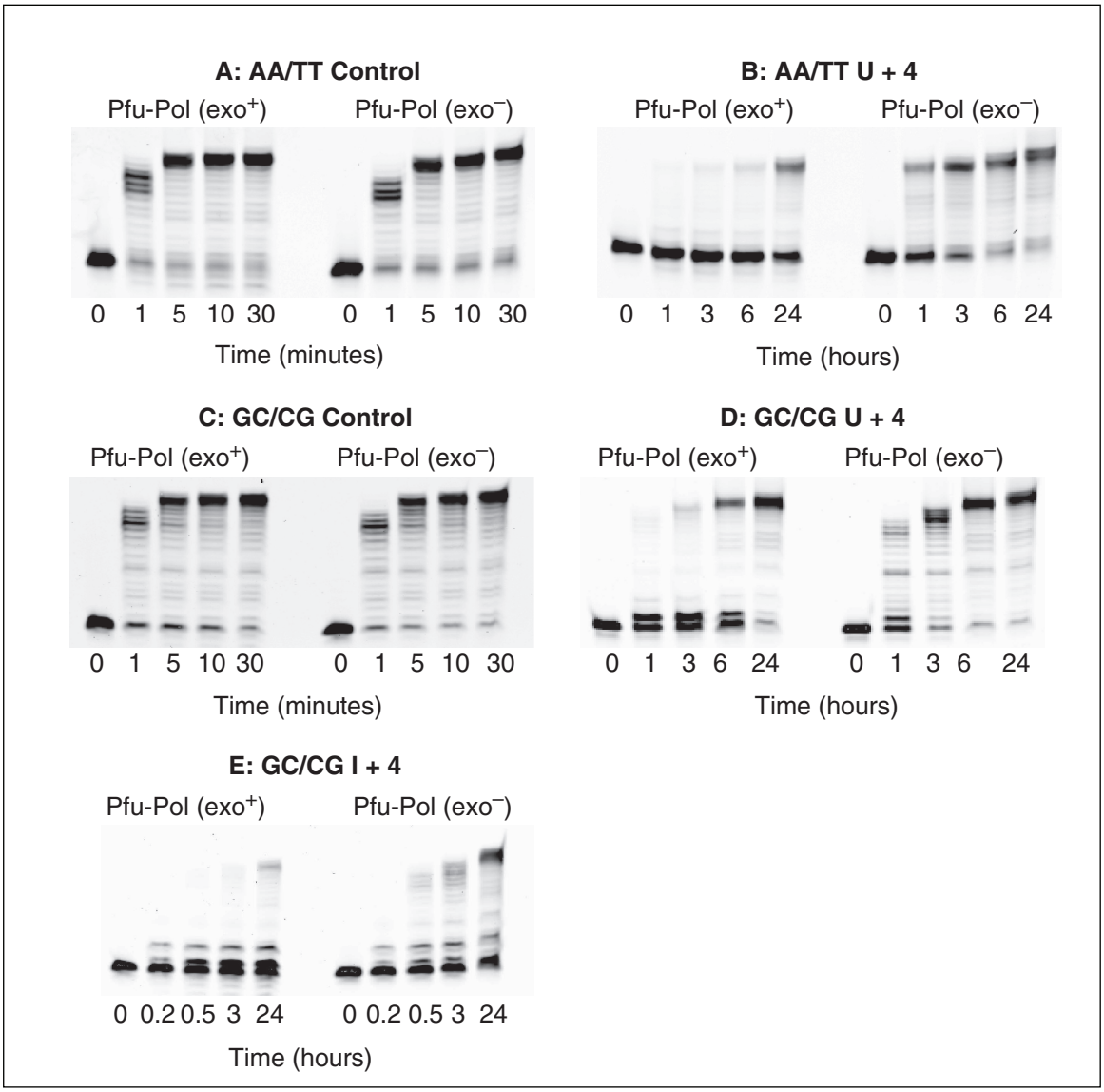


Figure 3.6: Primer extension assays performed using wild type (exo+) and D215A (exo-) versions of Pfu-Pol B with five different primer-templates, the names of which are provided above each image. The sequences of all of the primer-templates are given in full in table 2.5.

Reference to figure 3.6 reveals that, as previously observed, the profiles of extension for exo+ and exo- versions of Pfu-Pol B are almost identical with control primer-templates (**A** and **C**), except for the lack of exonucleolytic degradation at the bottom

of the images. Comparison of the U+4 and I+4 extensions (**B**, **D** and **E)**, however, reveals that the exo- strain of Pfu-Pol B is capable of bypassing both uracil and hypoxanthine much more readily than the exo+ strain, as observed in figure 3.5. With the exo- U+4 and I+4 assays a significant amount of fully extended product is visible after 1 hour, however with the wild type enzyme the majority of the starting material is still present after 6 hours. These results add further support to the role of 3'-5' exonucleolytic activity in deaminated base stalling.

3.5: Investigating the role of the uracil-binding pocket in inducing 3'-5' exonucleolytic activity

In order to investigate the role of the uracil-binding pocket in producing the stimulation of exonucleolytic activity observed in chapters 3.3 and 3.4, a number of the extension and exonuclease assays were repeated using exo+ and exo- strains of Pfu-Pol B in a V93Q mutant background. The V93Q mutation has previously been shown to completely eliminate the enzyme's ability to bind deaminated bases through a combination of steric exclusion from the uracil-binding pocket and elimination of hydrogen bonding (Fogg *et al.*, 2002; Firbank *et al.*, 2007). The gel images from these additional experiments are presented in figure 3.7.

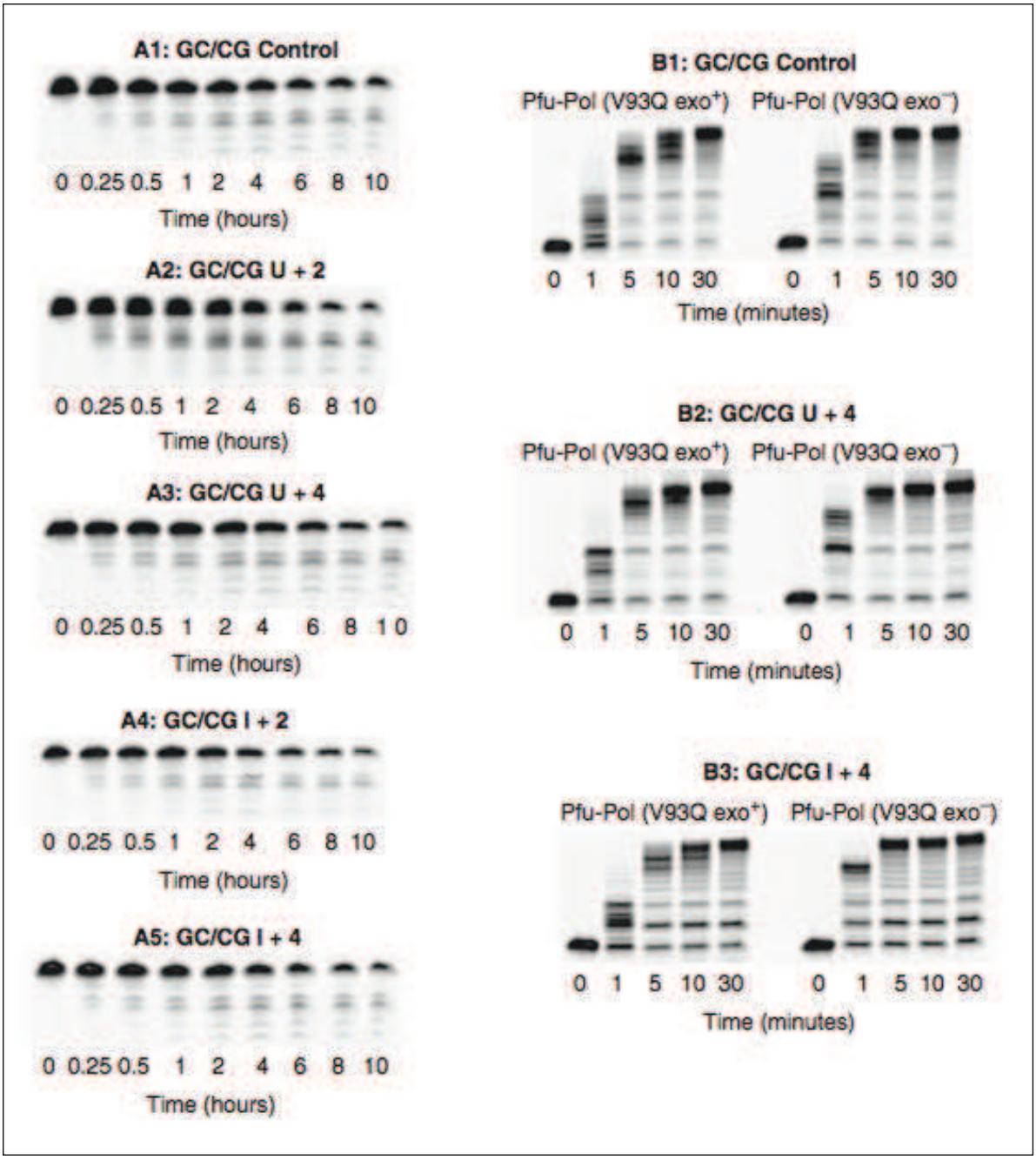


Figure 3.7: Exonuclease assays (**A1-A5**) and primer extension assays (**B1-B3**) performed using V93Q mutant versions of Pfu-Pol B. The extension reactions were performed using V93Q single and V93Q/D215A double mutant versions of Pfu-Pol B. Five different primer-templates were employed, the names of which are provided above each image. The sequences of each of the primer-templates are given in tables 3.1 and 3.3.

Reference to figure 3.7 reveals that using the V93Q mutant background completely eliminates the influence of the 3'-5' exonucleolytic in processing deaminated bases that was previously observed. All of the hypoxanthine- and uracil-containing primer-templates produce the same profiles of degradation and extension as the control substrates. The stimulatory effect of the bases on 3'-5' exonucleolytic activity when uracil and hypoxanthine are situated fewer than four bases from the primer-template junction is no longer observable. Similarly, the slight decrease in 3'-5' exonucleolytic activity when uracil and hypoxanthine are situated four bases ahead of the primer-template junction is also lost. The profiles of extension products are almost identical for exo+ and exo- variants with each of the uracil and hypoxanthine-containing primer-templates. The only minor difference being the slightly reduced rate of polymerisation observed for the V93Q exo+ enzyme, compared to the V93Q exo- variant.

Confirmation of these results comes from table 3.5, which reveals that all of the primer-templates produce similar rates of exonucleolysis. A slight 2.6-fold increase in exonucleolysis is seen with the U+2 primer-template, compared to the control, however, this may be attributed to residual binding by the pocket. It is also noted that the V93Q mutant has slightly reduced exonucleolytic activity with the control primer-template, compared to the wild type enzyme. At present the reasons for this remain unclear.

Sample	Repeat	Rate Constant (min^{-1})	Mean Rate Constant (min^{-1})	Standard Deviation	Relative rate
Control (gc/cg)	1 2 3	0.0027 0.0019 0.0031	0.0026	0.0006	1
U+2 (gc/cg)	1 2 3	0.0056 0.0056 0.0090	0.0067	0.0020	2.6
U+4 (gc/cg)	1 2 3	0.0035 0.0024 0.0044	0.0034	0.0010	1.3
I+2 (gc/cg)	1 2 3	0.0030 0.0032 0.0027	0.0030	0.0003	1.1
I+4 (gc/cg)	1 2 3	0.0028 0.0027 0.0027	0.0027	0.0001	1

Table 3.5: A summary of the single turnover rate constants calculated for the 3'-5' exonucleolytic activity of Pfu-Pol B V93Q with each of six tested substrates. The sequences of the substrates are all provided in table 3.3.

3.6: Discussion

From the results presented in chapter 3, one can conclude that stimulation of 3'-5' exonucleolytic activity by archaeal family B polymerases occurs in response to template strand deaminated bases. The stimulatory effect is seen when the deaminated base is situated between the 0 and +3 positions from the primer-template junction. The function of this stimulation appears to be to maintain the polymerase at its preferred position, 4 bases upstream from the deaminated base. Mismatched primer-templates resulted in an increase in 3'-5' exonucleolytic activity of roughly one order of magnitude, which is fairly typical for replicative polymerases (Johnson, 1993; Donlin *et al.*, 1991; Khare and Eckert, 2001). Interestingly, the stimulation in exonuclease activity in response to template strand deaminated bases (roughly 50-200 fold) is significantly greater than that seen with mismatched substrates. This is in spite of the fact that the duplexed regions of the U+2 and I+2 primer-templates contained only *bona fide* Watson-Crick base pairs. This apparent "urgency" of the polymerase to avoid replicative bypass of uracil and hypoxanthine is perhaps

indicative of the adverse biological consequences for the cell. Where uracil has arisen from spontaneous cytosine deamination, it may be excised from double-stranded DNA, but the repair pathway typically results in adenine incorporation opposite (Otterlei, 2000). As such, a c>t/g>a transition mutation would still occur at loci where spontaneous cytosine deamination has occurred.

The highest observed increase in exonuclease activity was with the U+2 (gc/cg) primer-template and was calculated to be 163 times higher than the rate for the relevant control. This was closely followed by the I+2 (gc/cg) primer-template, which produced an increase of 136-fold. Comparison of the absolute rates of exonucleolytic activity for the aa/tt and gc/cg primer-templates reveals that the rates for the aa/tt substrates were all significantly higher than those of the corresponding gc/cg substrates. For example, the rate constant for the control aa/tt primer-template was roughly 75-fold higher than that of the control gc/cg substrate. Similar ratios were found with other comparable pairs. The reason for this huge difference in rates can be attributed to the base pairs located at the 3' end of the primer. a/t base pairs share just two hydrogen bonds, as opposed to the three formed between g/c base pairs. As such, a/t base pairs form weaker interactions and are much more susceptible to thermo-dynamic influences and DNA "breathing" (Feng *et al.*, 1991). Since the stimulation in exonucleolytic activity we observed likely results from adoption of an editing conformation by the polymerase, DNA breathing would be expected to facilitate such a change (Marquez and Reha-Krantz, 1996). Based on previous studies, polymerase editing conformations appear to involve partial denaturation of the primer-template duplex at the 3' end of the primer. This denaturation facilitates the transposition of the primer terminus into the exonuclease domain (Freemont *et al.*, 1988; Baker and Bell, 1998). The faster rates of exonucleolysis observed with aa/tt primer-templates are therefore most likely due to the relative ease with which the primer-template duplex is unwound. The reason for the huge increase in 3'-5' exonucleolytic activity that was observed with the gc/cg U+2 and I+2 primer-templates may similarly be attributed to the lower rate of residual exonuclease activity that occurs with gc/cg primer-templates.

The U+4 and I+4 primer-templates (both aa/tt and gc/cg) were the only substrates that produced a decrease in exonuclease activity compared to their respective controls; the

mean k_{exo} for the U+4 aa/tt primer-template was roughly half that of the control, while the U+4 and I+4 gc/cg primer-templates showed a 20 % and 30 % decrease, respectively. It has previously been shown that polymerising activity is significantly inhibited when archaeal Pol B is bound 4 bases upstream from template strand deaminated bases (Greagg *et al.*, 1999; Fogg *et al.*, 2002; Shuttleworth *et al.*, 2004). The crystal structure discussed in chapter 3.1 of Tgo-Pol B bound to a U+4 substrate appears to indicate that the polymerase adopts an editing conformation with such substrates, however, the 3' terminal base in the primer strand is *not* located within the exonuclease active site. A primer-template bound in such a configuration would be acted upon poorly by both exonuclease and polymerising activities, thus explaining the observed results. It would therefore appear that, upon stalling, archaeal family B polymerases adopt a dormant state, in which both polymerising and 3'-5' exonucleolytic activities are reduced.

At the time of performing this work, no structural data was available of an archaeal family B polymerase bound to a DNA substrate containing uracil or hypoxanthine at a position where stimulation of exonucleolytic activity occurs, i.e. 0, +1, +2 or +3, relative to the primer-template junction. We propose that, as with the U+4 complex shown in figure 3.1, an archaeal family B polymerase bound to such a substrate would adopt an editing conformation, however, the 3' terminus of the primer strand would additionally be unwound to produce a single-stranded region whose terminal base would reside within the 3'-5' exonuclease active site. Although archaeal family B polymerases bind most tightly to DNA substrates with uracil at the +4 position, significant affinity is still observed with substrates containing uracil at the +1, +2 and +3 positions (Shuttleworth *et al.*, 2004). Previously, it was not clear how the enzyme maintained the interactions of the uracil-binding pocket with the deaminated base while the catalytic residues remained bound to the primer-template junction. Based on the results presented here, we propose that archaeal family B polymerases may induce an appropriate degree of melting of duplex regions in DNA primer-templates to re-position the primer-template junction within the catalytic site and maintain an effective separation of four bases between the polymerase active site and the uracil residue. It seems reasonable that the binding energy from the interaction with uracil would be sufficient to support the proposed strand partitioning behaviour of archaeal

family B polymerases. This partitioning behaviour appears to be consistent for both "strongly" (gc/cg) and "weakly" (aa/tt) base-paired primer-templates.

Another interesting observation from the results presented here is that V93Q exo+ Pfu-Pol B exhibits slightly reduced polymerising activity with all of the tested primer-templates, compared to the V93Q exo- enzyme, as shown in figure 3.7. The normal activity of archaeal family B polymerases relies on dynamic interplay between the polymerase and exonuclease domains of the enzyme, as well as the uracil-binding pocket and substrate-binding motifs. The V93Q mutation results in deformation of the uracil-binding pocket and hence null recognition of uracil and hypoxanthine. One possible explanation for the reduced polymerising activity of the V93Q exo+ enzyme is that the enzyme may be slightly less efficient at shuttling DNA substrates between the various domains.

In conclusion, the proposed model for the deaminated base response of archaeal family B polymerases can be described as follows. Upon encountering template strand uracil or hypoxanthine, the N-terminal uracil-binding pocket of archaeal Pol B traps the deaminated base, causing polymerisation to stall four bases upstream from the lesion. At this point, an "idling" behaviour is adopted, a diagrammatic representation of which is shown in figure 3.8. It should be noted that the stimulatory effect of deaminated bases is no longer observed when the residue is located at the -1 position. Once replicative bypass has been achieved, the polymerase can no longer specifically recognise and respond to template strand deaminated bases.

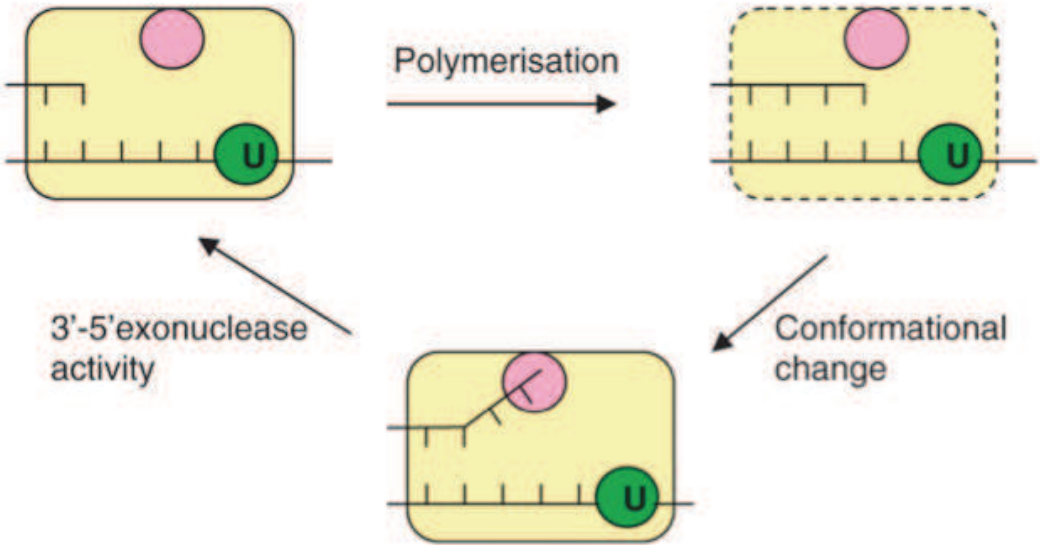


Figure 3.8: A simplified model of "idling", as is proposed to occur when archaeal Pol B encounters template strand uracil or hypoxanthine. In the first tile (top left) the polymerase, which is represented by the yellow box, is bound to a DNA primer-template with a template strand uracil residue situated at the preferred + 4 position. When adopting this conformation the enzyme enters a dormant mode of action with both polymerising and 3'-5' exonucleolytic activities inhibited. As nucleotides are incorporated, albeit very slowly, the enzyme-DNA complex enters a strained conformation, signified by the hatched yellow box in the top right of the figure. This strain induces a conformational change in the complex, which shifts the 3' terminus of the primer towards the 3'-5' exonuclease active site, which is indicated by the pink circle. 3'-5' exonucleolytic activity is then favoured and the primer-strand cleaved back to restore the 4 base separation of the primer-template junction and template strand deaminated base.

This model of idling is entirely compatible with previous studies, in which replicative polymerases have been shown to undergo repeated cycles of incorporating and removing standard nucleotides opposite template strand lesions (Johnson, 1993; Khare and Eckert, 2001; Khare and Eckert, 2002; Schwartz *et al.*, 1988; Garg *et al.*, 2004). In all known models of idling repeated dNTP to dNMP turnover is not accompanied by net DNA synthesis and, critically, replicative bypass of the damaged base is avoided. Interestingly, it has been suggested that accumulation of dNMP, or

conversely depletion of dNTP, may act as the primary signal for DNA repair pathways (Khare and Eckert, 2002). In previous studies, however, idling has always been observed to occur at mis-paired base pairs, which result from incorporation of a standard base opposite a template strand lesion. The resulting mispair is difficult for the polymerase to extend from and normally results in 3'-5' exonucleolytic degradation by the polymerase. The model of idling that we propose differs significantly in that the duplexed region of the primer-template contains only *bona fide* Watson-Crick base pairs and the lesion is located in the single-stranded region of the template.

The model proposed in figure 3.8 suggests that uracil or hypoxanthine situated at positions 0 to +3 promotes unwinding of the primer-template by archaeal Pol B, resulting in stimulation of 3'-5' exonucleolytic activity. Primer-template unwinding in response to deaminated bases has been further probed using the fluorescent base analogue 2-aminopurine (2AP). The results of these analyses are described in chapter 4. Additionally, a structure of Tgo Pol B bound to a primer-template mimic containing hypoxanthine at the +2 position has recently been published (Killelea *et al.*, 2010). This structure is discussed at the end of chapter 4.

Chapter 4

DNA duplex partitioning by archaeal family B polymerases upon encountering template strand deaminated bases

4.1: Background

As described in chapter 3, archaeal family B DNA polymerases appear to be capable of modulating the levels of polymerising and 3'-5' exonucleolytic activities they exhibit to prevent replicative bypass of template strand deaminated bases. As the enzyme approaches uracil or hypoxanthine, a stimulation in 3'-5' exonucleolytic activity of up to two orders of magnitude is observed when the enzyme is situated between 0 and 3 bases upstream from the lesion (Russell *et al.*, 2009). It is thought that this behavioural adjustment serves to prevent the replicative bypass of uracil or hypoxanthine by maintaining the polymerase 4 bases upstream from the lesion. After observing this increase in 3'-5' exonucleolytic activity, the mechanism by which the stimulation occurs was probed. Numerous polymerases have been shown to adopt an editing conformation following nucleotide misincorporation (Freemont *et al.*, 1988; Baker and Bell, 1998; Datta *et al.*, 2009). Adoption of the editing conformation involves transposition of the 3' terminus of the primer from the polymerase active site to the 3'-5' exonuclease active site. This transposition represents a significant conformational change with the two sites typically separated by a distance of 30-40 Å (Freemont *et al.*, 1988; Wang *et al.*, 1997; Kamtekar *et al.*, 2004). By manipulating the primer terminus in this way, the polymerase is capable of promoting proofreading exonucleolytic activity and thus the likelihood of the misincorporated base being excised from the primer is stochastically favoured. Based on the results presented in chapter 3, as well as a recently-solved co-crystal structure of Tgo Pol B (Firbank *et al.*, 2008), it was proposed that archaeal family B polymerases might similarly undergo a conformational change to promote 3'-5' exonucleolytic activity, however, rather than occurring in response to base misincorporation, the proposed change would occur in response to template strand deaminated bases (Russell *et al.*, 2009).

In all known editing conformations the terminal 3-4 bases at the 3' end of the primer completely dissociate from the template strand and instead occupy the channel leading to the 3'-5' exonuclease domain (Shamoo and Steitz, 1999; Hogg *et al.*, 2004; Subuddhi *et al.*, 2008). It has been observed that, following base misincorporation, the polymerising activity of the Klenow fragment is inhibited up to 4 bases beyond the mismatched base pair (Miller and Grollman, 1997). In other words, even after

replicative bypass of a mismatch has occurred, the polymerase is capable of "remembering" and responding to the error. The same phenomenon has also been observed with several other polymerases, including HIV reverse transcriptase and DNA polymerase α (Johnson, 1993; Ng *et al.*, 1989). Structural studies of mismatched DNA have revealed that perturbation in the overall structure of DNA is detectable at primer-template junctions up to 6 bases beyond duplexed mismatches, thus providing an explanation for the short-term "memory" of these enzymes (Johnson and Beese, 2004). The editing conformation of archaeal Pol B that is proposed to arise in response to template strand deaminated bases differs from those previously described in that the duplexed region of the primer-template contains only *bona fide* Watson-Crick base pairs.

In addition to the crystallographic studies mentioned above, dynamic changes in DNA structures have been studied extensively using fluorescent nucleotide analogues, such as 2-aminopurine (2AP) (Guest *et al.*, 1991; Stivers, 1999; Jean and Hall, 2001; Hariharan and Reha-Krantz, 2005; Datta *et al.*, 2009). 2AP is an artificial analogue of adenine and guanine. It forms base pairs with thymine in a Watson-Crick geometry (Nordlund *et al.*, 1989; Law *et al.*, 1996) and with cytosine in a wobble configuration (Sowers *et al.*, 1986; Fagan *et al.*, 1996; Sowers *et al.*, 2000). Its unusual fluorescence properties make it ideally suited to real time investigation of local conformational changes in DNA. After excitation at a wavelength of 315 nm, 2AP exhibits a peak emission at around 370 nm (Bandwar and Patel, 2001; Liu and Martin, 2002). The intensity of this emission is dependent on the degree of stacking of the 2AP residue such that an unstacked base fluoresces with much greater intensity than a stacked one (Jean and Hall, 2001). By analysing the fluorescence state of 2AP in DNA, it is therefore possible to determine the nature of the environment in which it resides, particularly in terms of the stacking interactions that it forms.

In this chapter a combination of steady state and time-resolved 2AP fluorescence analyses are described in order to demonstrate the stacking states of 2AP residues located near the 3' ends of DNA duplexes. Targeted mutagenesis of regions proposed to play a role in modulating 3'-5' exonucleolytic activity by partitioning DNA strands will also be presented, together with the *in vitro* results of these modifications on proofreading exonuclease activity.

4.2: Optimisation of conditions for fluorescence analyses

4.2.1: 2AP-containing DNA substrates

In order to test the theory of primer-template strand separation being the primary mechanism by which stimulation of 3'-5' exonucleolytic activity is achieved when archaeal Pol B responds to template strand deaminated bases, the artificial base analogue 2AP was employed. A series of primer-templates was synthesised containing 2AP at either the final or penultimate position at the 3' end of the primer strand (table 4.1). Template strands were fully complementary to the primer within the duplexed region and contained uracil at either the +2 or +4 position, relative to the primer-template junction. Control templates containing only the four canonical bases were also synthesised. Two alternative ^{2A}PA U+4 primer-templates were used in analyses, with one containing guanine at the +2 position, relative to the primer-template junction, and the other containing thymine. The reasons for the two alternative configurations are described in chapters 4.3 and 4.4.

Name	Sequence
ss $A^{2A}P$	5′-(Cy5-)GGGGATCCTCTAGAGTCGACCTGCAGGGCAP-3′
$A^{2A}P$ control (g)	5′-(Cy5-)GGGGATCCTCTAGAGTCGACCTGCAGGGCAP-3′ 3′-CCCCTAGGAGATCTCAGCTGGACGTCCCGTTCGTTCGAACAGAGG-5′
$A^{2A}P$ U+2	5′-(Cy5-)GGGGATCCTCTAGAGTCGACCTGCAGGGCAP-3′ 3′-CCCCTAGGAGATCTCAGCTGGACGTCCCGTTCUTTCGAACAGAGG-5′
$A^{2A}P$ U+4 (g)	5′-(Cy5-)GGGGATCCTCTAGAGTCGACCTGCAGGGCAP-3′ 3′-CCCCTAGGAGATCTCAGCTGGACGTCCCGTTCGTUCGAACAGAGG-5′
ss ^{2A}PA	5′-(Cy5-)GGGGATCCTCTAGAGTCGACCTGCAGGGCPA-3′
^{2A}PA control (g)	5′-(Cy5-)GGGGATCCTCTAGAGTCGACCTGCAGGGCPA-3′ 3′-CCCCTAGGAGATCTCAGCTGGACGTCCCGTTCGTTCGAACAGAGG-5′
^{2A}PA U+2	5′-(Cy5-)GGGGATCCTCTAGAGTCGACCTGCAGGGCPA-3′ 3′-CCCCTAGGAGATCTCAGCTGGACGTCCCGTTCUTTCGAACAGAGG-5′
^{2A}PA U+4 (g)	5′-(Cy5-)GGGGATCCTCTAGAGTCGACCTGCAGGGCPA-3′ 3′-CCCCTAGGAGATCTCAGCTGGACGTCCCGTTCGTUCGAACAGAGG-5′
^{2A}PA U+4	5′-(Cy5-)GGGGATCCTCTAGAGTCGACCTGCAGGGCPA-3′ 3′-CCCCTAGGAGATCTCAGCTGGACGTCCCGTTCTTUCGAACAGAGG-5′

Table 4.1: Names and sequences of primer-templates used in fluorescence analyses and 3'-5' exonuclease assays in this chapter. 2AP is indicated by an underlined green P. Uracil is indicated by an underlined red U. The residue at the +2 position in the ^{2A}PA U+4 primer-template, "G", was later switched to "T" for reasons discussed in chapters 4.2.4 and 4.Y. Both residues are underlined and highlighted in blue. Cy5-labelled substrates were used to perform gel shift and 3'-5' exonuclease assays. Fluorescence analyses were performed using the unlabelled primer-templates.

4.2.2: Buffer compositions

A D215A mutant version of Pfu-Pol B was used in most of the fluorescence analyses presented in chapter 4.2 to prevent exonucleolytic degradation of substrates (Evans *et al.*, 2000). Fluorescence measurements were attempted using wild type Pfu-Pol B with calcium substituted for magnesium in the reaction buffer, since Ca^{2+} ions cannot normally replace Mg^{2+} in the 3'-5' exonucleolytic pathway (Cowan, 1998). However, trace levels of degradation were always observed; presumably due to minute

quantities of contaminating magnesium or manganese ions. A third buffer was therefore employed for analyses of the wild type enzyme, containing EDTA in place of the divalent cation component. The EDTA served to chelate any divalent cations, thereby inhibiting proofreading exonucleolytic activity. The three buffers used in fluorescence analyses in this chapter are fully described in table 4.2, together with a fourth buffer that was used for K_D determination by fluorescence anisotropy in chapter 4.5.

Buffer name	Buffer composition
Mg	20 mM Tris-HCl (pH 8.5), 10mM KCl, 2 mM $MgCl_2$, 10 mM $(NH_4)_2SO_4$
Ca	20 mM Tris-HCl (pH 8.5), 10mM KCl, 2 mM $CaCl_2$
EDTA	20 mM Tris-HCl (pH 8.5), 10mM KCl, 2 mM EDTA, 10 mM $(NH_4)_2SO_4$
HEPES	10 mM HEPES-NaOH (pH 7.5), 100 mM NaCl, 1 mM EDTA

Table 4.2: Names and compositions of the buffers used in fluorescence studies and binding assays for K_D determination in chapter 4 of this thesis.

4.2.3: Primer-template annealing and saturation by Pfu-Pol B

Based on previous studies, 2AP-containing primer-templates were employed at a final concentration of 2 μM to produce a sufficiently strong fluorescence signal for detection (Su *et al.*, 2004; Lee and Berdis, 2006). A 2-fold excess of template over primer was used to ensure that the 2AP-containing primer strand was fully bound, since single-stranded 2AP would be expected to greatly increase the fluorescence emission of the duplexes (Rist and Marino, 2002). Mobility shift assays were performed to confirm annealing of the substrates. Assessment of the annealing of the ^{2A}PA control (g), ^{2A}PA U+2 and ^{2A}PA U+4 (g) primer-templates is shown in figure 4.1, indicating that > 95% of the primer was bound by the template strand with each of the duplexes. Virtually identical results were obtained with the A^{2A}P primer-templates.

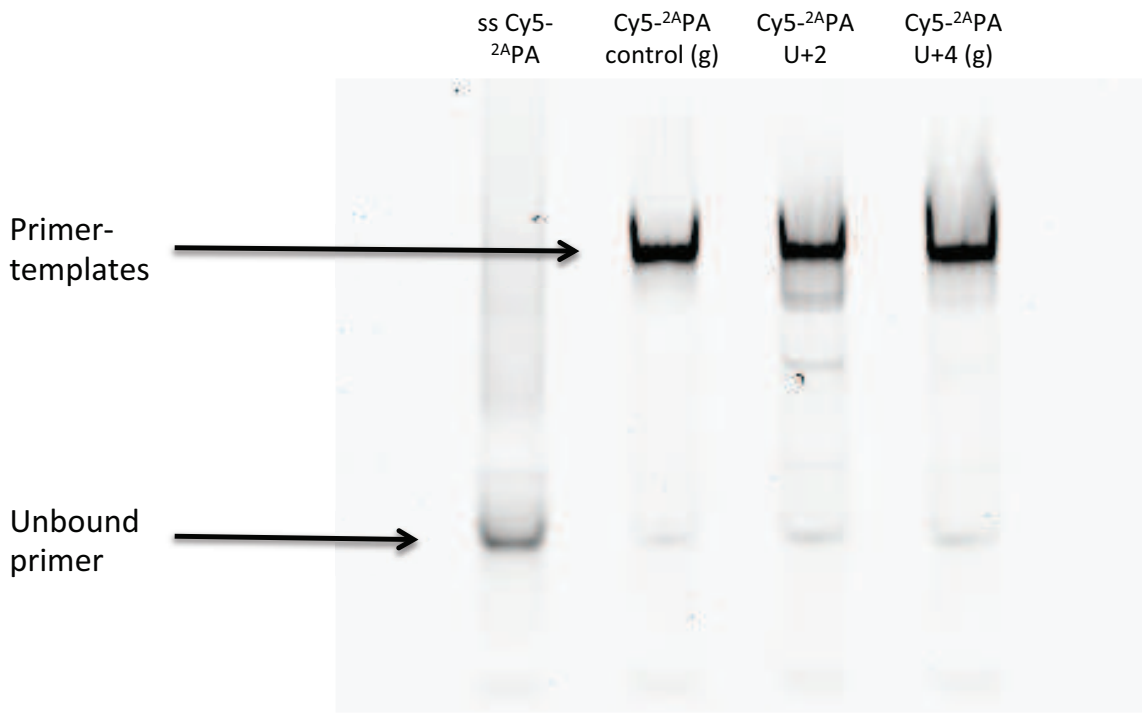


Figure 4.1: Mobility shift assays, performed by native PAGE analysis, showing the migration of the single-stranded ^{2A}PA DNA, together with the ^{2A}PA control (g), ^{2A}PA U+2 and ^{2A}PA U+4 (g) primer-templates (described in table 4.1). The primer and duplexes were diluted in Mg buffer (described in table 4.2).

Further mobility shift assays were performed to establish the concentration of enzyme required to saturate the DNA under experimental conditions. Primer-templates were incubated with a range of concentrations of Pfu-Pol B (D215A) in Mg buffer (table 4.2). An image from one such assay, showing the migration of the ^{2A}PA U+2 and ^{2A}PA U+4 (g) primer-templates is shown in figure 4.2.

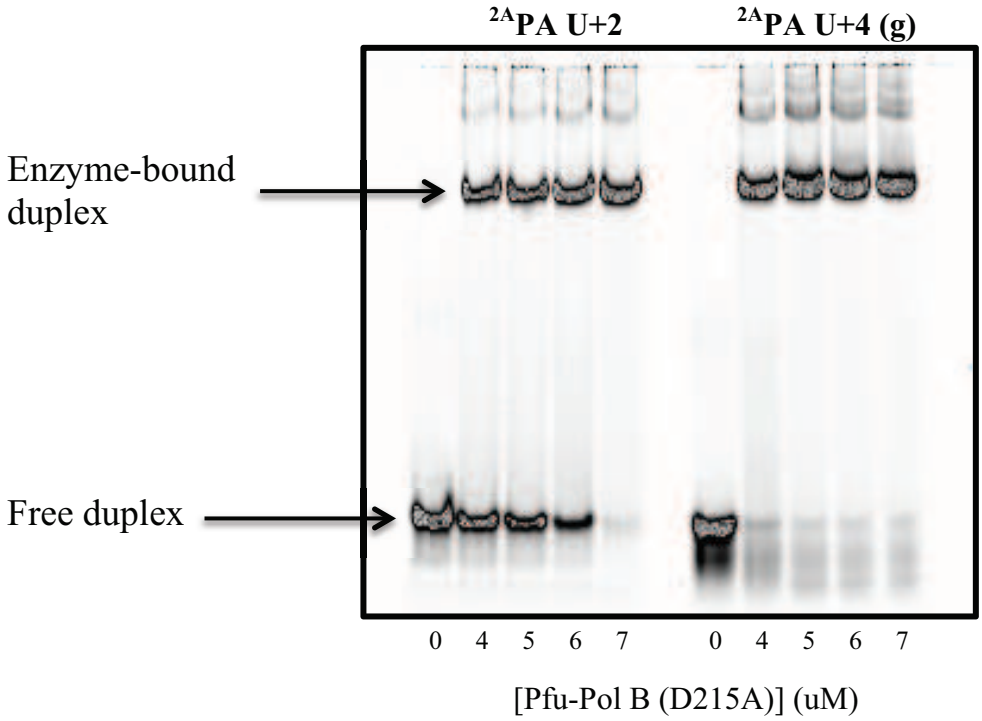


Figure 4.2: Mobility shift assays establishing the concentration at which 2AP-containing primer-templates were fully bound by Pfu-Pol B (D215A) in Mg buffer (table 4.2). The substrates were incubated with concentrations of enzyme ranging from 0-7 μM before analysis by native PAGE. The sequences of the primer-templates are provided in table 4.2.

Figure 4.2 reveals that the ^{2A}PA U+2 and ^{2A}PA U+4 (g) primer-templates were both fully bound using 7 μM enzyme. In agreement with previous studies, the binding affinity for the U+4 primer-template was significantly higher than for the U+2 (Shuttleworth *et al.*, 2004). Mobility shift assays were not performed using the other buffers or the wild type enzyme, however, the binding data presented in chapter 4.5 confirm that exonuclease-proficient and -deficient versions of Pfu-Pol B bind DNA substrates with the same affinity, regardless of the metal ion composition of the buffer. As such, it was assumed that the data in figure 4.2 was representative of all possible conditions. Unfortunately, similar assays performed using the control primer-templates revealed that, even with a 10-fold excess of enzyme, saturation could not be achieved due to the lower affinity of Pfu-Pol B for DNA substrates that contain only the four canonical bases (Shuttleworth *et al.*, 2004; Gill *et al.*, 2007). In order to increase affinity, primer-template-enzyme-PCNA complexes were also tested (data not shown), however, the background fluorescence of PCNA exacerbated

measurements and a considerable excess of enzyme was still required. It was therefore decided that fluorescence analyses would be performed using only the U+2 and U+4 primer-templates. For analyses of DNA-enzyme complexes, a final concentration of 8 μM enzyme was employed in all subsequent fluorescence analyses.

4.2.4: Preliminary steady state fluorescence measurements

Preliminary steady state fluorescence measurements were made using both the $A^{2A}P$ and ^{2A}PA sets of primer-templates, i.e. with the 2AP residue situated at either the terminal or penultimate position at the 3' end of the primer. After excitation at 315 nm, the fluorescence emission values of the unbound duplexes and the enzyme-DNA complexes were recorded at a wavelength of 370 nm. The fluorescence emission values are summarised in table 4.3.

Primer-template name	Duplex fluorescence at 370 nm (AU)	Complex fluorescence at 370 nm (AU)
$A^{2A}P$ U+2	529	1659
$A^{2A}P$ U+4 (g)	635	1152
^{2A}PA U+2	180	1387
^{2A}PA U+4 (g)	196	552

Table 4.3: Fluorescence emission values, after excitation at 315 nm, of the ^{2A}PA and $A^{2A}P$ U+2 and U+4 primer-templates described in table 4.1. The fluorescence emission values were recorded at 370 nm in Mg buffer (table 4.2) for the unbound primer-templates and for the duplexes in complex with Pfu-Pol B (D215A).

Initial results revealed that both of the U+2 primer-templates exhibited a significant increase in fluorescence emission at 370 nm upon binding by the enzyme, indicating a substantial degree of strand separation. A much smaller increase was observed with the U+4 primer-templates. When measured in isolation, however, the fluorescence emission values of both of the $A^{2A}P$ primer-templates were significantly higher than those of the ^{2A}PA substrates (529 AU and 635 AU for $A^{2A}P$, compared to 180 AU and

196 AU for ^{2A}PA). As such, the relative increase in fluorescence for the A^{2A}P primer-templates, upon binding by the enzyme, was reduced. The differing fluorescence states of the unbound primer-templates can be attributed to the stacking interactions formed by the 2AP residues. When situated at the penultimate position the 2AP residue possesses an extra base on its 3' end and, as such, its fluorescence is partially quenched by the additional stacking interactions that it forms (Stivers, 1998; Jean and Hall. 2001). On the basis of the data presented in table 4.3, it was therefore decided that the ^{2A}PA primer-templates would be used in all subsequent fluorescence analyses in order to achieve the best possible signal to noise ratio.

In order to gain a more accurate measure of the fluorescence states of the primer-templates under different conditions, fluorescence emission spectra were recorded between 340 nm and 400 nm for the ^{2A}PA U+2 and ^{2A}PA U+4 (g) primer-templates. Again, measurements were made with the substrates alone and in complex with Pfu-Pol B (D215A). The resulting spectra can be seen in figure 4.3.

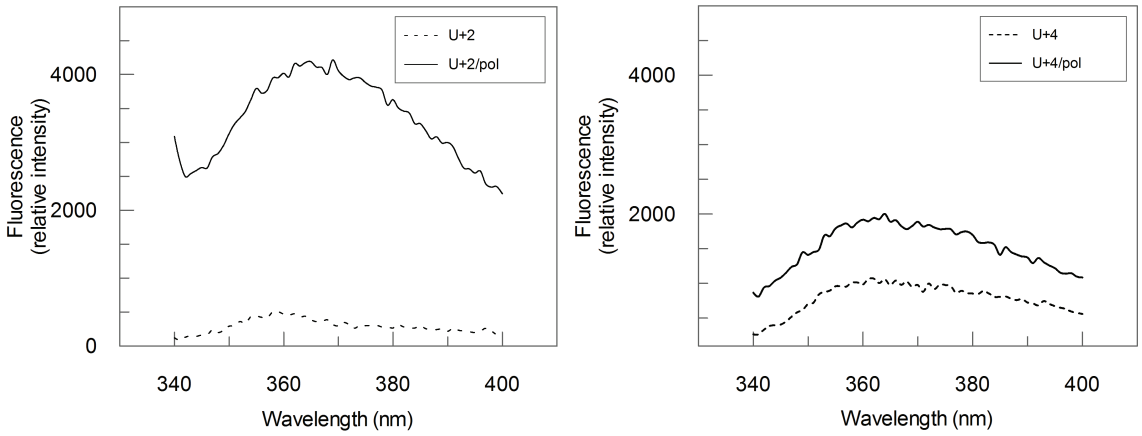


Figure 4.3: Fluorescence emission spectra of ^{2A}PA U+2 and ^{2A}PA U+4 (g) primer-templates alone and in complex with Pfu-Pol B D215A after excitation at 315 nm.

Figure 4.3 reveals that, upon binding by the enzyme, the fluorescence emission at 370 nm of the U+2 primer-template increased roughly 12-fold; from about 300 AU to 4000 AU, indicating significant strand separation. The U+4 primer-template exhibited only a 2-fold increase. However, comparison of the spectra of the primer-templates on their own indicates that the unbound U+4 primer-template fluoresces with much greater intensity than the unbound U+2 primer-template. As such, the

observed increase upon binding by Pol B appears reduced. Based on previous studies, it is known that 2AP fluorescence is context-specific and somewhat sensitive to the nature of the surrounding bases (Holz and Weinhold, 1997). The sequences of the templates were therefore compared, revealing that the ^{2A}PA U+4 (g) template contained guanine at the +2 position, as opposed to the uracil residue found in the U+2 strand. The U+4 template was therefore redesigned to contain thymine at this position, which is much more similar in nature to uracil (table 4.1). Repetition of the measurements shown in figure 4.3 using the new ^{2A}PA U+4 primer-template, containing thymine at +2, resulted in significantly less fluorescence emission for the unbound U+4 substrate (figure 4.4). Significantly, the emission spectra for both unbound primer-templates became virtually indistinguishable, while the emission spectrum for the U+4-enzyme complex remained largely the same as with ^{2A}PA U+4 (g). The new U+4 primer-template now exhibited an increase in fluorescence emission at 370 nm of approximately 5-fold upon binding by the enzyme, primarily due to the decrease in unbound fluorescence. The new ^{2A}PA U+4 primer-template was therefore used in all subsequent analyses, instead of ^{2A}PA U+4 (g).

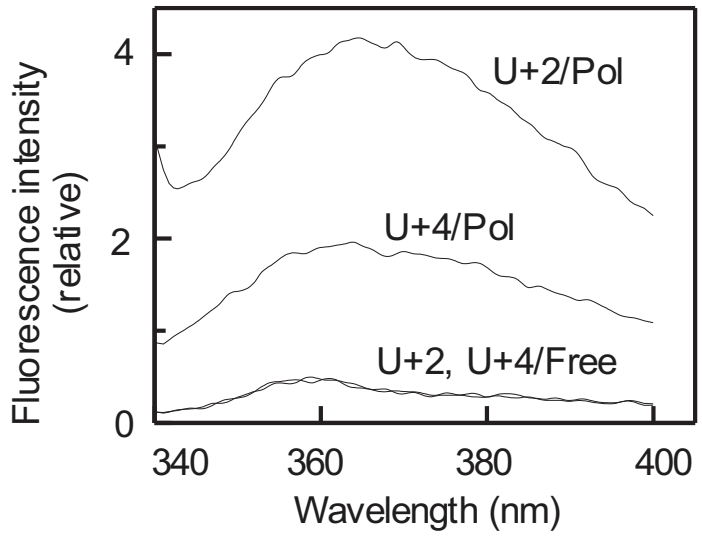


Figure 4.4: Fluorescence emission spectra of ^{2A}PA U+2 and ^{2A}PA U+4 primer-templates (labelled "U+2" and "U+4") after excitation at 315 nm. The emission spectra of the DNA substrates alone are labelled "/Free", while those of the enzyme-primer-template complexes are labelled "/Pol".

4.3: Full analysis of steady state 2AP fluorescence using optimised conditions

In order to improve the accuracy of the fluorescence data, spectra were recorded for the Mg, Ca and EDTA buffers (described in table 4.2) alone. The spectra were virtually indistinguishable from one another and revealed a distinctive raman band between 340 nm and 370 nm (figure 4.5). The spectra of the buffers were then subtracted from measurements of the unbound primer-templates to produce the corrected data shown in figure 4.7. The fluorescence emission spectra of the enzymes were also recorded in each of the buffers (figure 4.6) and subsequently subtracted from the spectra of the complexes to produce the data shown in figure 4.7.

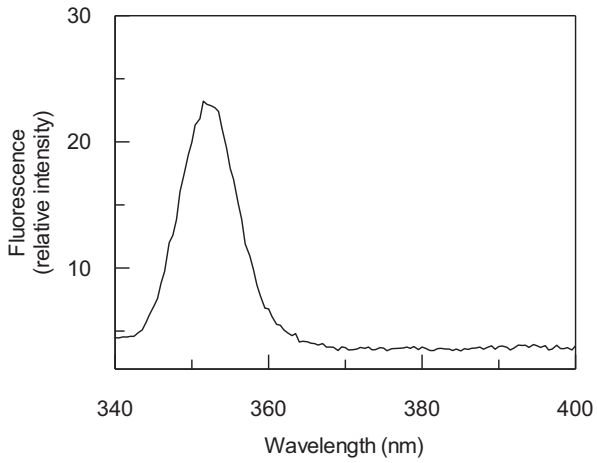


Figure 4.5: Fluorescence emission spectrum of the EDTA buffer (described in table 4.2) after excitation at 315 nm. The distinctive shape of the Raman band is visible between 340 nm and 370 nm. Virtually identical spectra were seen with the Mg and Ca buffers.

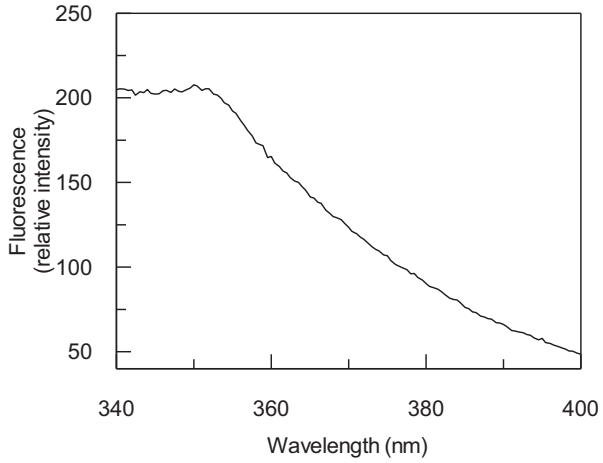


Figure 4.6: Fluorescence emission spectrum of Pfu-Pol B exo- in the EDTA buffer (described in table 4.2) after excitation at 315 nm. Virtually identical spectra were seen with the Mg and Ca buffers and with the exo+ enzyme in EDTA buffer.

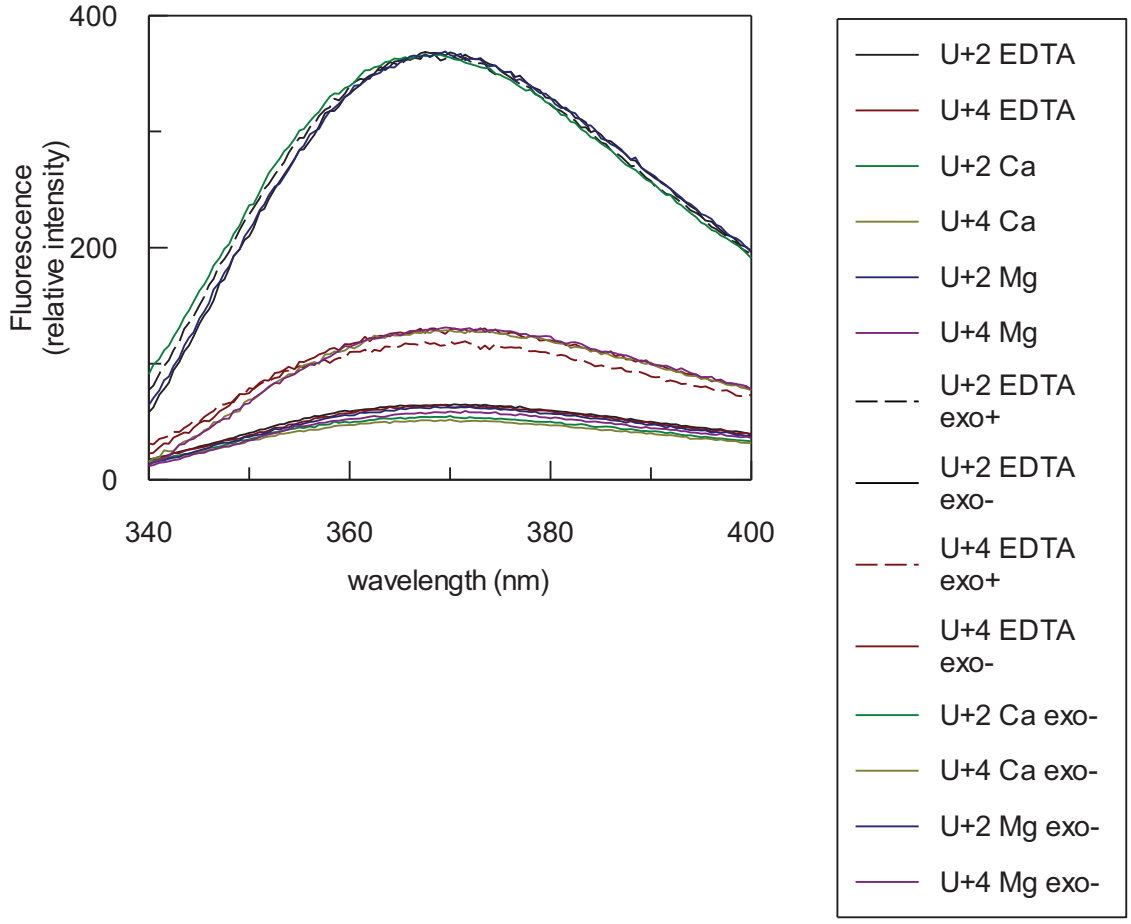


Figure 4.7: Corrected fluorescence emission spectra of the ^{2A}PA U+2 and ^{2A}PA U+4 primer-templates described in table 4.1. Spectra were recorded for the substrates on their own and in complex with Pfu-Pol B (exo+ and exo-). Three different buffers were employed, each of which is described fully in table 4.2.

The preliminary spectra shown in figures 4.3 and 4.4 exhibited a peak in fluorescence emission around 360 nm. Figure 4.7 reveals that, after correction, all of the spectra exhibited peak fluorescence emission at around 370 nm, which agrees well with previous studies (Bandwar and Patel, 2001; Liu and Martin, 2002). After correction, the spectra of the two unbound primer-templates are essentially indistinguishable from one another, regardless of the buffer they were recorded in. Upon binding by both the wild type and exo- enzymes, the U+4 primer-template exhibited a small but significant increase in fluorescence emission, indicating a small degree of strand partitioning. Again, the buffer composition appeared to make little difference to the U+4 spectra. Upon binding by both the wild type and exo- enzymes, the U+2 primer-template exhibited a much larger increase in fluorescence emission, indicating a significant degree of strand separation. Once again, the buffer composition appeared

to make little difference to the spectra. The greater increase in fluorescence emission of the U+2 primer-template would appear to indicate that its 2AP residue is experiencing less stacking interactions than that of the U+4 primer-template, upon binding by the enzyme, and that it enters a single-stranded context. The lower increase in fluorescence of the U+4 primer-template suggests that its 2AP residue remains predominantly stacked and in a double-stranded context.

4.4: Time-resolved measurement of fluorescence decay curves of 2AP-containing primer-templates and enzyme complexes

All of the data presented in chapter 4.4 were produced in Edinburgh by Dr Anita Jones, Dr David Dryden and Ms Xiaohua Wu, as part of a collaboration.

In order to further elucidate the nature of the conformational changes that follow binding of uracil-containing primer-templates by archaeal Pol B, use was made of time-resolved fluorescence. The form of the 2AP fluorescence decay function is an excellent reporter of local environment and, in favourable instances, can provide specific structural information e.g. in monitoring base flipping by DNA methyltransferases (Neely *et al.*, 2005; Youngblood *et al.*, 2008). Time-resolved 2AP fluorescence has also been used to investigate viral family-B DNA polymerases using 2AP in both primer (Subuddhi *et al.*, 2008) and template strands (Hariharan and Reha-Krantz, 2005). The experiments described in this section were all carried out with D215A (exo-) Pfu-Pol B to eliminate difficulties arising from the slow hydrolysis observed with the wild type enzyme. The decay parameters for the free U+2 and U+4 primer-templates are given in table 4.4 and, in graphical form, in figure 4.8. It should be noted that the U+4 primer-templates contained thymine at the +2 position.

Solution composition	τ_1/ ns	τ_2/ ns	τ_3/ ns	τ_4/ ns	A_1	A_2	A_3	A_4
U+2 / Ca^{2+}	0.04	0.64	2.8	8.1	0.80	0.11	0.07	0.02
U+2 / Mg^{2+}	0.05	0.63	2.7	8.1	0.77	0.12	0.08	0.03
U+2 / Ca^{2+} / Pfu-Pol B[a]	0.17	0.94	3.6	9.0	0.42	0.29	0.19	0.10
U+2 / Mg^{2+} / Pfu-Pol B[a]	0.17	0.99	3.7	8.7	0.40	0.23	0.23	0.14
U+4 / Ca^{2+}	0.04	0.46	2.5	8.2	0.88	0.06	0.04	0.02
U+4 / Mg^{2+}	0.04	0.44	2.4	8.2	0.89	0.06	0.03	0.02
U+4 / Ca^{2+} / Pfu-Pol B[a]	0.06	0.67	3.6	9.5	0.64	0.12	0.15	0.09
U+4 / Mg^{2+} / Pfu-Pol B[a]	0.06	0.67	3.5	9.4	0.66	0.11	0.15	0.08

[a]In all cases the Pfu-Pol B used in these experiments lacked 3′-5′ proofreading exonuclease activity (D215A).

Table 4.4. Fluorescence lifetimes (τ_i) and their fractional amplitudes (A_i) for the 2AP-containing primer templates ^{2A}PA U+2 and ^{2A}PA U+4 (see table 4.1) in the absence and presence of Pfu-Pol B (D215A). In all cases measurements were carried out with buffers containing either Ca^{2+} or Mg^{2+} as the divalent cation (table 4.2).

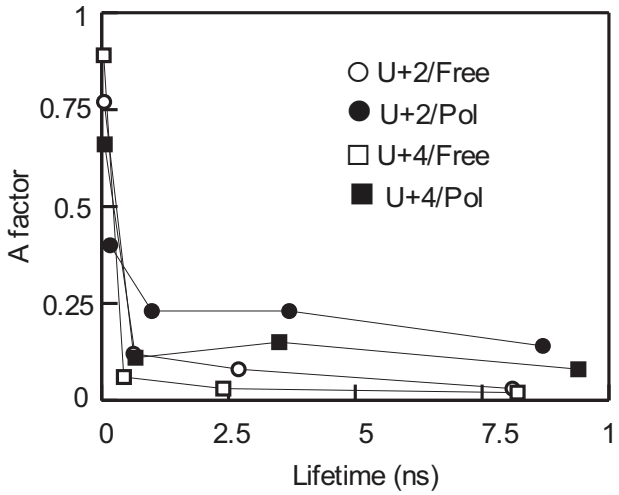


Figure 4.8: Graphical representation of the 2AP fluorescence decay parameters for the free and polymerase-bound ^{2A}PA U+2 and ^{2A}PA U+4 primer-templates (the sequences of which are provided in table 4.1). The excitation wavelength used was 315 nm and the concentration of primer-templates was 2 μM. Polymerase, when added, was at a concentration of 8 μM. The fluorescence decays were recorded in the Mg buffer (table 4.2) using a D215A (exo-) variant of Pfu-Pol B. The plot shows the fractional amplitude (A factor) versus lifetime.

The two primer-templates have essentially identical properties and require four lifetime (τ) components to give a satisfactory fit. The four-component decay, commonly observed for 2AP in DNA (Jean and Hall, 2001; Neely *et al.*, 2005; Youngblood *et al.*, 2008), shows that 2AP experiences a heterogeneous environment and indicates the existence of at least four conformational states of the duplex. The A factors indicate the fraction of the emitting 2AP population with a given lifetime and hence the fractional population of each conformational state. One conformation has a very short lifetime of 50 ps (τ_1) and is dominant, accounting for ~80% of the duplex population (A_1). This corresponds to a highly stacked structure in which excited 2AP is rapidly quenched by charge transfer interaction with neighbouring bases. The longest decay time (8 ns) is similar to that of free 2AP-riboside (Neely *et al.*, 2004) and is attributed to 2AP in an extrahelical environment, free from interbase quenching. This species constitutes a very small fraction of the population, ~3 %. The two intermediate decay times are due to imperfectly stacked conformations in which 2AP is intrahelical, but quenching is much less efficient than in the closely stacked structure. Together the intermediate conformations account for ~20 % of duplexes. The results shown in figure 4.8 were obtained in the presence of Mg^{2+}. Very similar data were obtained when Mg^{2+} was substituted with Ca^{2+} (not shown).

Substantial changes were observed when Pfu-Pol exo- was added to the U+2 primer-template (table 4.4, figure 4.8). All lifetimes increased, indicating a loss of quenching and a change in local environment. Of particular significance is the large increase in τ_1 to 170 ps and the considerable decrease in A_1 to 0.4. The greater value of τ_1 indicates a smaller degree of stacking in this conformational state and the decrease in A_1 (and concomitant increases in A_2 to A_4) shows a large transfer of duplex population to more poorly stacked states. These effects are entirely consistent with unwinding of the duplex as seen structurally with primer-templates containing hypoxanthine at +2 and discussed in greater detail in chapter 4.7 (Killelea *et al.*, 2010). As found with the steady-state measurements, much less perturbation was observed for U+4 in complex with the polymerase. Noticeably the effect on the shortest decay component, highly stacked 2AP, was markedly different than for U+2. A very short decay time (60 ps) persists indicating the continued existence of a highly stacked state, which remained the dominant conformation, ~65 % of the population (table 4.4,

figure 4.8). The value of τ_2 also remains significantly shorter than in the U+2 complex, again indicating the persistence of more well stacked conformations. The values of τ_3 and τ_4 are similar in both complexes, indicating that 2AP in poorly stacked or extrahelical states experiences a similar environment in both cases, as might be expected. Overall the decay parameters for U+4 bound to the polymerase indicate a modest distortion of the duplex, resulting in some transfer of population from well-stacked conformations (τ_1 and τ_2) to poorly stacked or extrahelical states, but with the vast majority of the population, ~76 % (A_1 + A_2), remaining in the well stacked states. Again these data agree well with a polymerase-U+4 structure, where no major structural changes in the duplex region of the primer-template are observed (Firbank *et al.*, 2008).

Time-resolved fluorescence analyses also provide further support for the findings of chapter 4.2.4; that guanine, as opposed to thymine, at +2 in the ^{2A}PA U+4 primer-template results in increased fluorescence emission for the unbound primer-template. The fluorescence lifetimes and fractional amplitudes of the two ^{2A}PA U+4 primer-templates and enzyme complexes are compared in tables 4.5 and 4.6 and in graphical form in figures 4.9 and 4.10. The decay parameters of the unbound ^{2A}PA U+4 duplex (with thymine at +2) are much more similar to those of unbound ^{2A}PA U+2, compared to those of the ^{2A}PA U+4 (g) primer-template.

Solution composition	τ_1/ ns	τ_2/ ns	τ_3/ ns	τ_4/ ns	A_1	A_2	A_3	A_4
U+4 (g) / Ca^{2+}	0.09	0.68	2.7	7.7	0.51	0.26	0.17	0.06
U+4 (g) / Mg^{2+}	0.09	0.67	2.7	7.7	0.54	0.25	0.16	0.05
U+4 (g) / Ca^{2+}/Exo-	0.11	0.91	3.8	9.2	0.49	0.18	0.22	0.11
U+4 (g) / Mg^{2+}/Exo-	0.11	1.0	3.9	9.2	0.46	0.16	0.26	0.12

Table 4.5: Fluorescence lifetimes (τ_i) and their fractional amplitudes (A_i) for the ^{2A}PA U+4 (g) duplex and its complex with Pfu-Pol B (D215A). Data, collected at 3 emission wavelengths, were analysed globally to give the reported lifetimes. The fractional amplitudes (A factors) show little variation with emission wavelength; those for 365-nm emission are given. The average lifetimes over all emission wavelengths (<τ>) are also given.

Solution composition	τ_1/ ns	τ_2/ ns	τ_3/ ns	τ_4/ ns	A_1	A_2	A_3	A_4	$<\tau>$/ns
U+4 (t) / Ca^{2+}	0.04	0.46	2.5	8.2	0.88	0.06	0.04	0.02	<0.33
U+4 (t) / Mg^{2+}	0.04	0.44	2.4	8.2	0.89	0.06	0.03	0.02	<0.33
U+4 (t) / Ca^{2+}/Exo-	0.06	0.67	3.6	9.5	0.64	0.12	0.15	0.09	1.8
U+4 (t) / Mg^{2+}/Exo-	0.06	0.67	3.5	9.4	0.66	0.11	0.15	0.08	1.6

Table 4.6: Fluorescence lifetimes (τ_i) and their fractional amplitudes (A_i) for the ^{2A}PA U+4 duplex and its complex with Pfu-Pol B (D215A). Data, collected at 3 emission wavelengths, were analysed globally to give the reported lifetimes. The fractional amplitudes (A factors) show little variation with emission wavelength; those for 365-nm emission are given. The average lifetimes over all emission wavelengths ($<\tau>$) are also given.

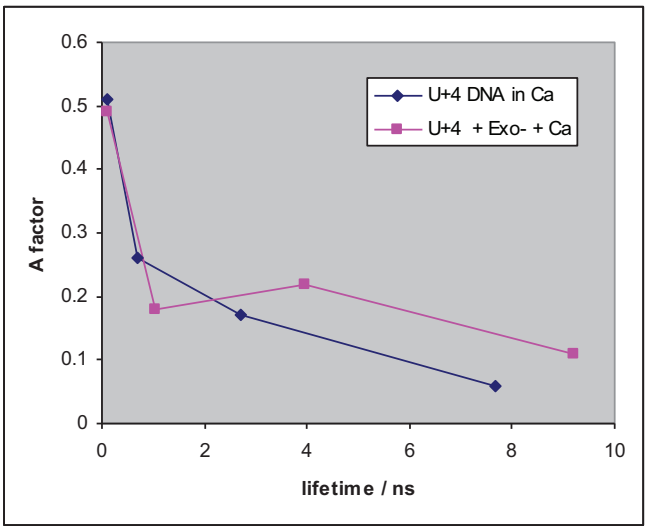


Figure 4.9: Plot of A factor versus lifetime for ^{2A}PA U+4 (g) in Ca buffer in the absence and presence of Pfu-Pol B (D215A).

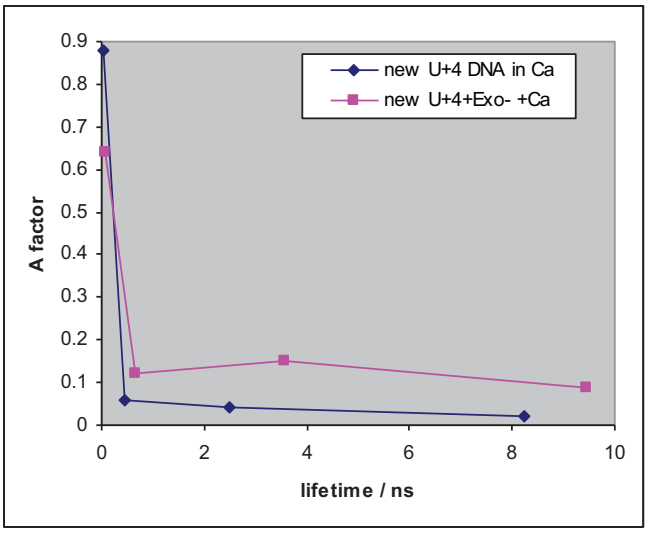


Figure 4.10: Plot of A factor versus lifetime for ^{2A}PA U+4 in Ca buffer in the absence and presence of Pfu-Pol B (D215A).

4.5: Probing the role of a ß-hairpin loop region of archaeal Pol B in DNA binding and modulation of DNA strand separation

In order to gain a greater understanding as to which physical features of archaeal Pol B might be facilitating the strand partitioning behaviour observed in chapters 4.3 and 4.4, reference was made to a co-crystal structure of *Thermococcus gorgonarius* family B DNA polymerase (Firbank *et al.*, 2008) bound to a U+4 primer-template mimic. The image (shown in figure 4.11) revealed a ß-hairpin motif that forms part of the exonuclease domain. The motif is made up of two antiparallel ß strands joined by a loop and, based on studies of the RB69 and T4 DNA polymerases, critically influences the choice between primer strand extension or degradation (Shamoo and Steitz, 1999; Reha-Krantz, 2010; Trzenecka *et al.*, 2009; Subuddhi *et al.*, 2008). Interestingly, the ß-hairpin appears to rest directly on the primer-template junction when Tgo-Pol B is bound with uracil at the +4 position. It was proposed that as archaeal Pol B incorporates dNTPs into the primer strand (albeit at a reduced rate),

the ß-hairpin might insert between the two strands, shifting the primer towards the exonuclease domain to promote degradation, rather than extension.

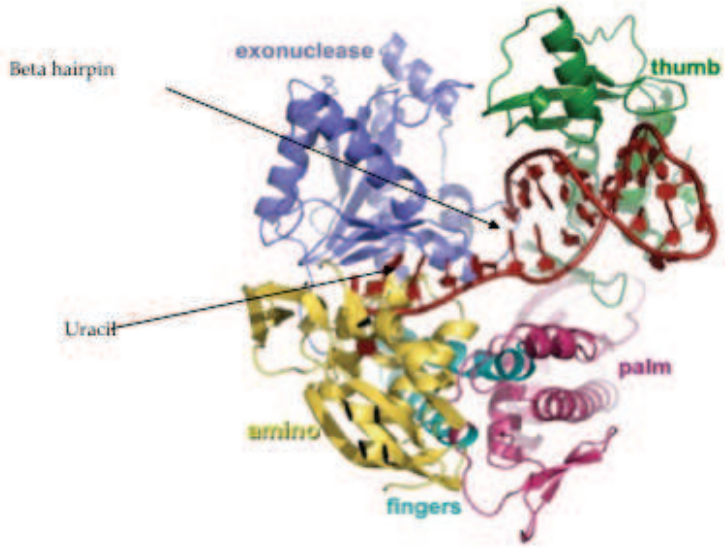


Figure 4.11: Co-crystal structure of the family B DNA polymerase of *Thermococcus gorgonarius* in complex with a DNA hairpin designed to simulate a primer-template with uracil at the +4 position, relative to the primer-template junction (Firbank *et al.*, 2008). The bound uracil residue is labelled on the left of the figure, together with a beta hairpin region that is proposed to play a role in modulating 3'-5' exonucleolytic activity by partitioning the primer and template strands of bound DNA substrates.

Further support for the beta hairpin region of archaeal Pol B playing a role in strand partitioning was provided by a more recent co-crystal structure of Tgo-Pol B bound to a primer-template mimic with hypoxanthine at the +2 position (Killelea, 2010). The structure appears to show a separation of the primer and template strands at the 3' end of the primer, with the terminal two residues of the primer strand almost completely dissociated from the template (figure 4.12). On examining the structure more closely, two residues in particular, arginine 247 and tyrosine 261, were identified as potentially playing an important role in the strand separation process (Firbank *et al.*, 2008; Killelea *et al.*, 2010). The residues have been highlighted and can be seen resting on the template strand of the bound DNA substrate in figure 4.12. The corresponding residues in Pfu-Pol B, methionine 247 and tyrosine 261, were therefore substituted for alanine by site-directed mutagenesis. The mutations were introduced

on their own and in combination with one another, both in a wild type and D215A (exo-) mutant background. The 6 mutant enzymes were then purified and subjected to binding titrations, using fluorescence anisotropy to calculate their K_Ds with a variety of substrates and to establish that the modifications had not compromised the enzymes' ability to bind uracil-containing DNA.

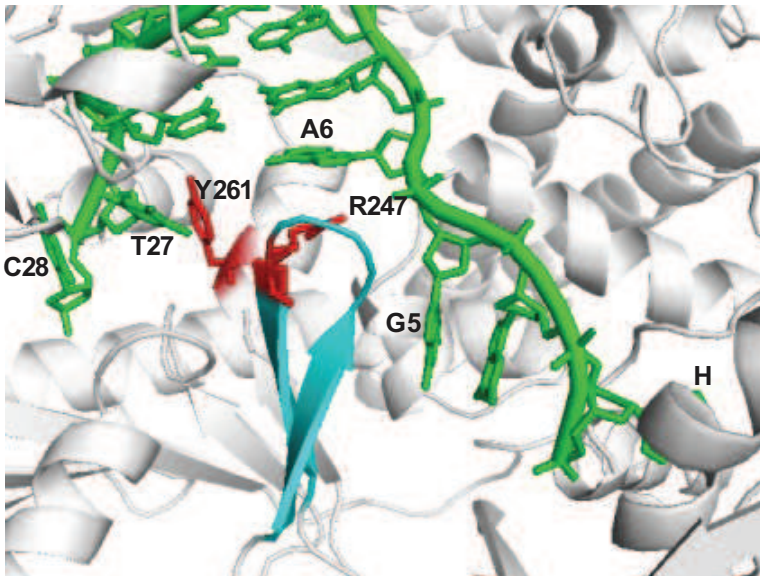


Figure 4.12: A close up image highlighting the amino acid residues of *Thermococcus gorgonarius* Pol B that form contacts with the primer-template junction of a bound DNA substrate containing hypoxanthine at +2 (shown in green). The amino acid residues, tyrosine 261 and arginine 247, were identified as potentially playing a critical role in strand separation and have been highlighted in red and cyan, respectively (Killelea *et al.*, 2010).

The DNA substrates listed in table 4.7 were used to determine the dissociation constants of all 8 wild type and mutant enzymes. To enable comparisons to be made with previously published data (Shuttleworth *et al.*, 2004), an alternative HEPES buffer was employed (table 4.2). The binding affinities of the enzymes were also measured in the EDTA buffer used for 2AP fluorescence experiments (table 4.2). The summarised dissociation constants are provided in tables 4.8 and 4.9. As seen in chapter 4.2.3, the affinity of Pfu-Pol B for the U+4 primer-template was significantly higher than for U+2, agreeing well with previously published data (Shuttleworth *et al.*,

2004). The binding curves of the enzymes are displayed in figure 4.13 and, for any given substrate/buffer combination, are essentially indistinguishable from one another.

Name	Sequence
ss	5'-Hex-TTTCTGGTUCCAGCTGGACCATTCGCCTATAGGACCTATT-3'
U+2	5'-Hex-TTTCTGGTUCCAGCTGGACCATTCGCCTATAGGACCTATT-3' 3'-GTCGACCTGGTAAGCGGATATCCTGGATAA-5'
U+4	5'-Hex-TTTCTGGTUCCAGCTGGACCATTCGCCTATAGGACCTATT-3' 3'-CGACCTGGTAAGCGGATATCCTGGATAA-5'

Table 4.7: Names and sequences of the DNA substrates used to perform binding titrations in chapter 4.4. Uracil residues are indicated by a red U. Hexachlorofluorescein labels were employed to enable measurement of fluorescence anisotropy.

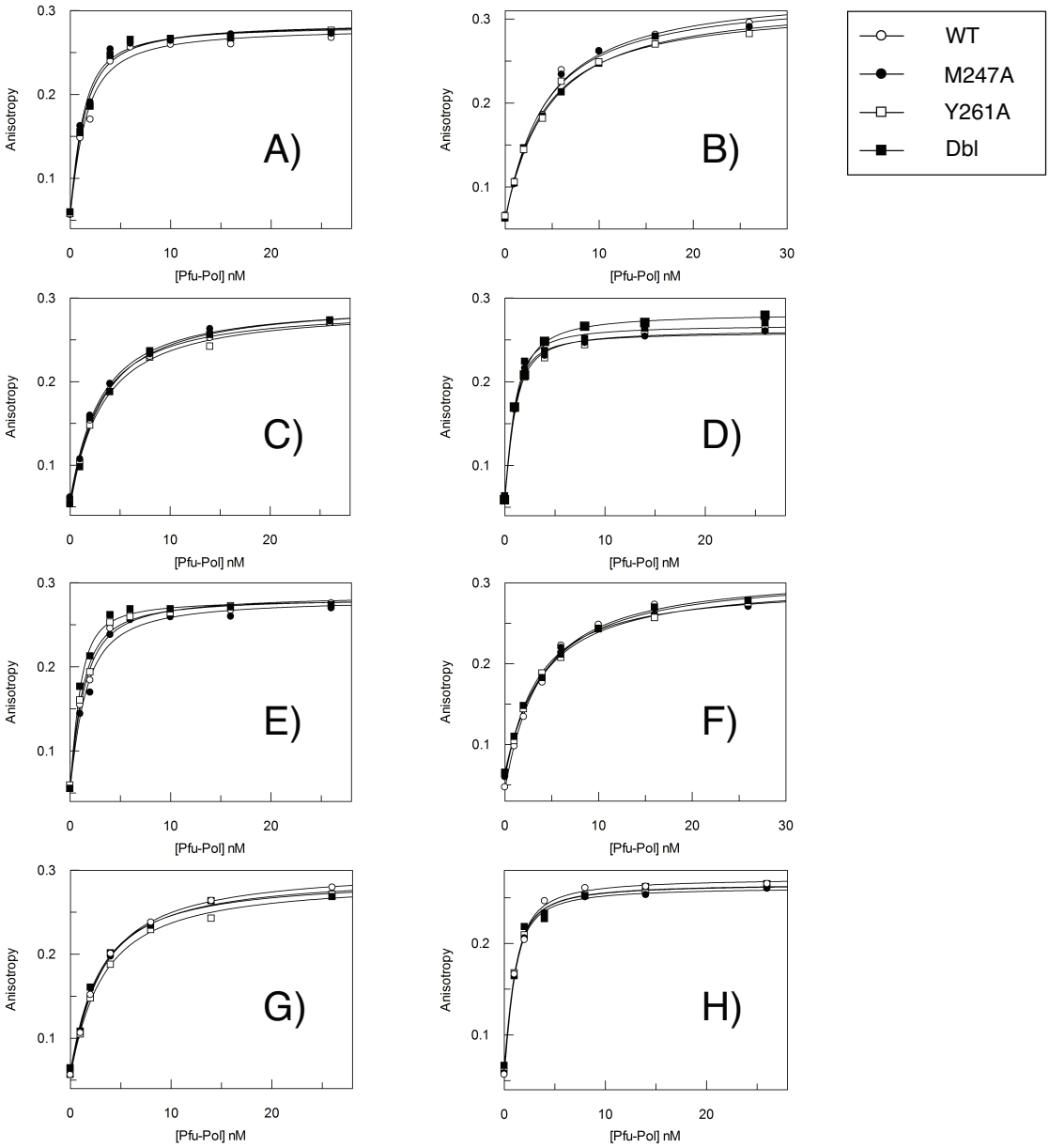


Figure 4.13: Binding curves of wild type and ß loop mutants of Pfu-Pol B with the DNA substrates listed in table 4.2, as determined by fluorescence anisotropy. **A)** Exo- enzymes with ss DNA in EDTA buffer. **B)** Exo- enzymes with ss DNA in HEPES buffer. **C)** Exo- enzymes with U+2 primer-template in HEPES buffer. **D)** Exo- enzymes with U+4 primer-template in HEPES buffer. **E)** Exo+ enzymes with ss DNA in EDTA buffer. **F)** Exo+ enzymes with ss DNA in HEPES buffer + EDTA. **G)** Exo+ enzymes with U+2 primer-template in HEPES buffer + EDTA. **H)** Exo+ enzymes with U+4 primer-template in HEPES buffer + EDTA.

DNA substrate and buffer composition	Calculated K_Ds (nM) of exo+ Pfu-Pol B variants			
	Wild type	M247A	Y261A	M247A/Y261A
ss in EDTA buffer	1.0	0.8	0.7	0.8
ss in HEPES buffer	3.5 ± 0.3*	3.5 ± 0.4*	3.5 ± 0.3*	3.4 ± 0.3*
U+2 in HEPES buffer	2.3	2.2	2.7	2.5
U+4 in HEPES buffer	0.6	0.5	0.4	0.4

Table 4.8: Summarised K_Ds of wild type and ß loop mutant variants of Pfu-Pol B exo- with the substrates listed in table 4.7, as determined by fluorescence anisotropy. The exact buffer compositions are described in table 4.2. *K_Ds for single-stranded DNA in the HEPES buffer represent the mean of three values ± the standard deviation.

DNA substrate and buffer composition	Calculated K_Ds (nM) of exo+ Pfu-Pol B variants			
	Wild type	M247A	Y261A	M247A/Y261A
ss in EDTA buffer	0.9	0.7	1.0	0.4
ss in HEPES buffer	3.0 ± 0.2*	3.2 ± 0.4*	3.3 ± 0.3*	3.3 ± 0.4*
U+2 in HEPES buffer	2.5	2.4	2.6	2.3
U+4 in HEPES buffer	0.5	0.5	0.5	0.5

Table 4.9: Summarised K_Ds of wild type and ß loop mutant variants of Pfu-Pol B exo+ with the substrates listed in table 4.7, as determined by fluorescence anisotropy. The exact buffer compositions are described in table 4.2. *K_Ds for single-stranded DNA in the HEPES buffer represent the mean of three values ± the standard deviation.

After confirming that the beta loop mutants retained the ability to bind uracil-containing substrates with the same affinity as wild type Pfu-Pol B, all 8 enzymes were used to measure fluorescence emission spectra with the ^{2A}PA U+2 and ^{2A}PA U+4 primer-templates previously described. Again, the buffer and enzyme backgrounds were subtracted from the spectra to produce the data shown in figures 4.14-4.17. With all of the tested buffers and in both the wild type and exo- backgrounds, the ß loop mutant complexes exhibited significantly reduced fluorescence emission, compared to the corresponding wild type values, providing strong evidence for the role of M247 and Y261 in strand partitioning. Interestingly,

with the EDTA buffer the spectra of the ß loop mutant-U+4 complexes were virtually indistinguishable from those of the unbound primer-templates, indicating that in the absence of a divalent cation the mutants were completely incapable of partitioning the U+4 primer-template. The fluorescence emission of all of the U+2 complexes (wild type and ß loop mutants) remain significantly higher than those of the corresponding U+4 values.

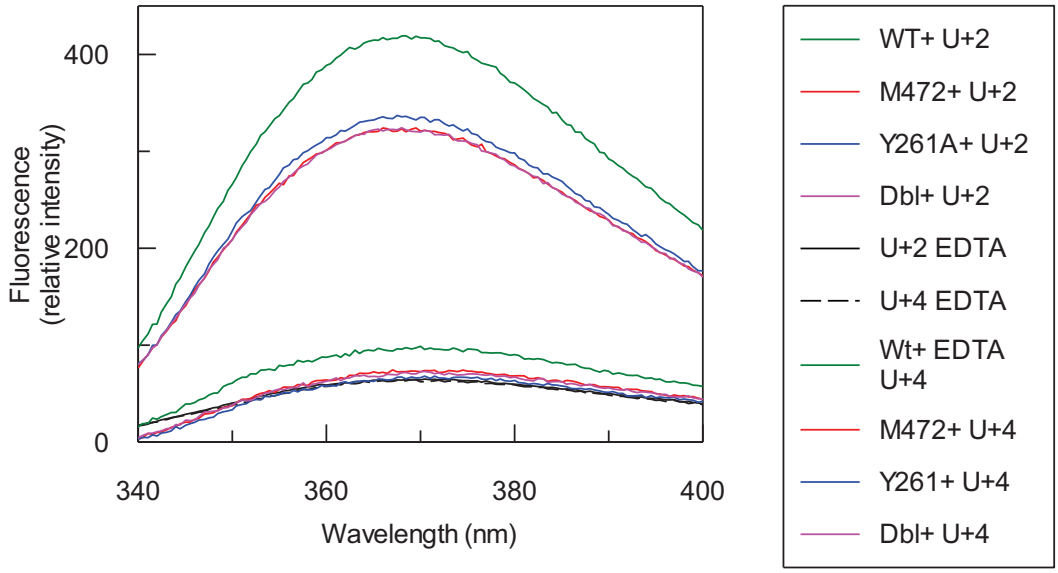


Figure 4.14: Corrected fluorescence emission spectra of Pfu-Pol B wild type and ß loop mutants in an exo+ background after excitation at 315 nm. Samples were suspended in the EDTA described in table 4.2 to prevent degradation of substrates.

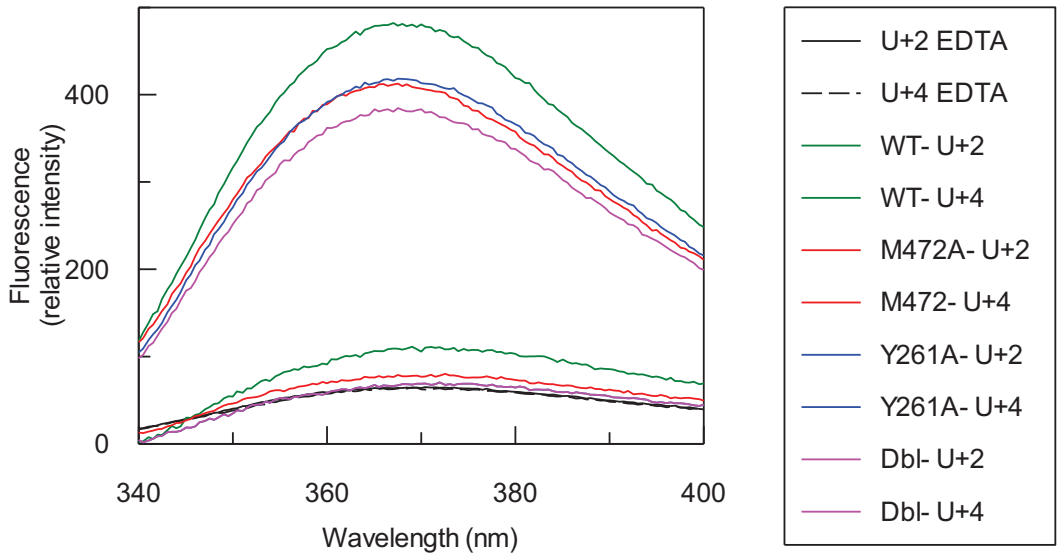


Figure 4.15: Corrected fluorescence emission spectra of Pfu-Pol B wild type and ß loop mutants in an exo- background after excitation at 315 nm. Samples were suspended in the EDTA buffer described in table 4.2.

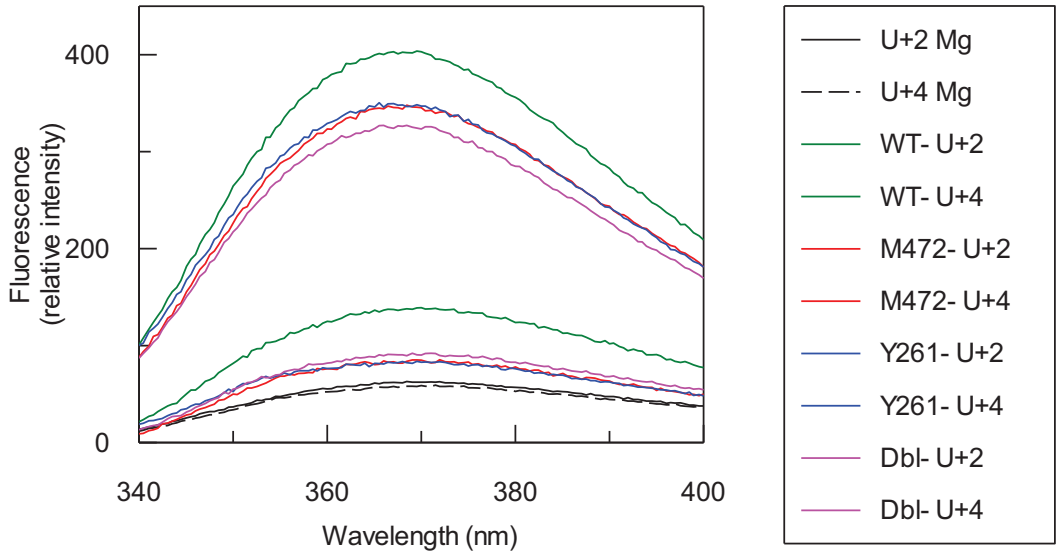


Figure 4.16: Corrected fluorescence emission spectra of Pfu-Pol B wild type and ß loop mutants in an exo- background after excitation at 315 nm. Samples were suspended in the Mg buffer described in table 4.2.

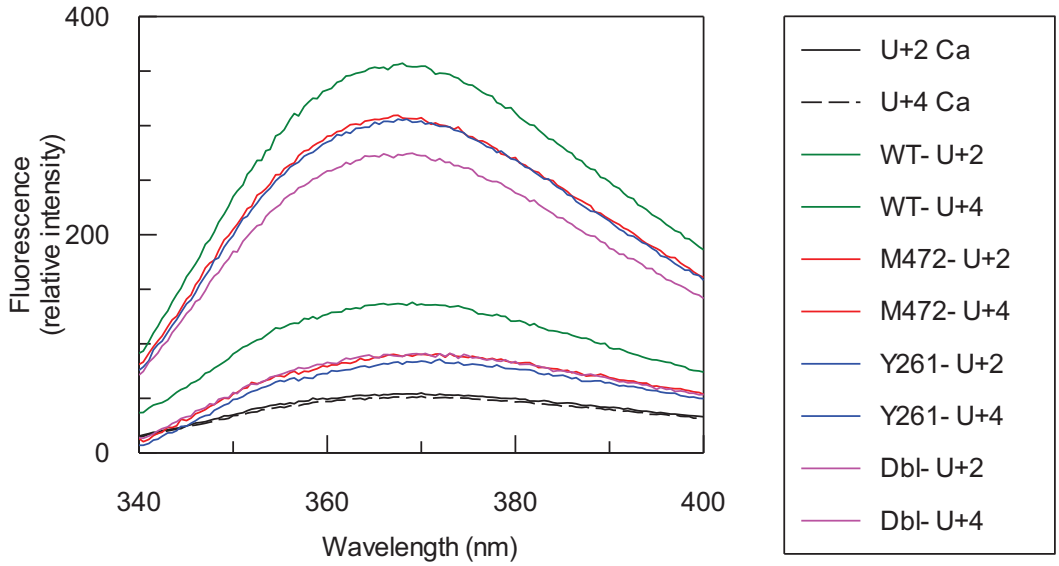


Figure 4.17: Corrected fluorescence emission spectra of Pfu-Pol B wild type and ß loop mutants in an exo- background after excitation at 315 nm. Samples were suspended in the Ca buffer described in table 4.2.

4.6: Probing the role of a ß-hairpin loop region of archaeal Pol B in modulating 3'-5' exonucleolytic activity

In order to gain a greater understanding of the role of the ß-hairpin loop in modulating proofreading exonucleolytic activity by archaeal Pol B, exonuclease assays were performed using the Cy5-labelled ^{2A}PA control, U+2 and U+4 primer-templates shown in table 4.1. The M247A, Y261A and M247A/Y261A double mutant enzymes in the wild type (exo+) background were used to perform the analyses. A selection of gel images produced with the fully wild type enzyme can be seen in figure 4.18. The calculated single turnover rates of 3'-5' exonucleolysis (k_{exo}s) of all of the enzymes are summarised in table 4.10.

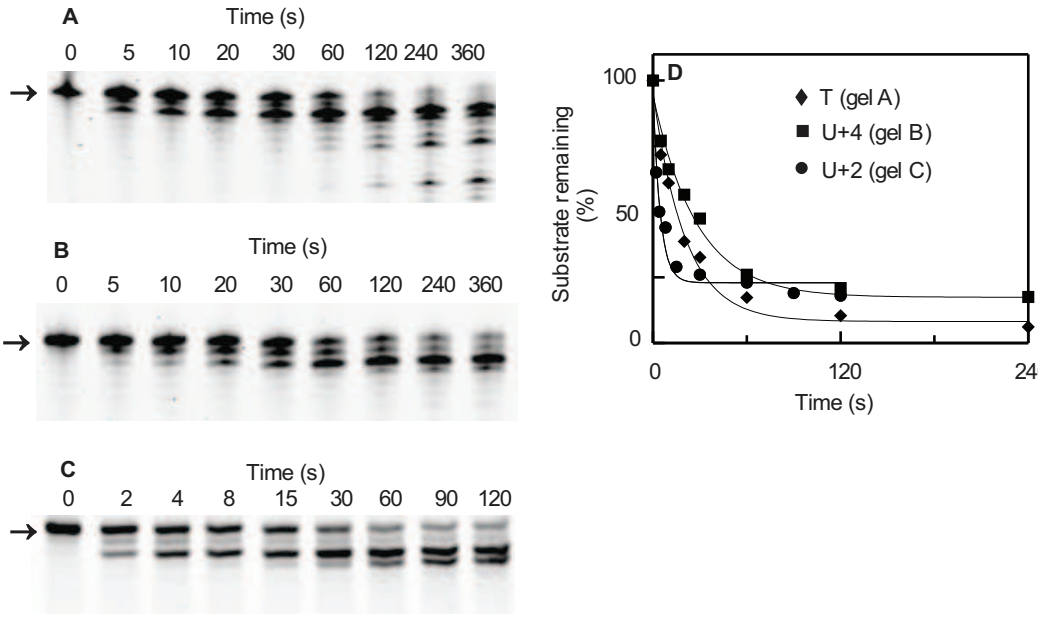


Figure 4.18: Hydrolysis of ^{2A}PA primer-templates (see table 4.1) by the proofreading exonuclease activity of wild type Pfu-Pol B. **A)** T (control) primer-template **B)** U+4 primer-template **C)** U+2 primer-template. Panels A-C show the degradation of the Cy5-labelled primer strand (starting material indicated by the arrows) with time, as assessed by denaturing PAGE. **D)** Data obtained in panels A-C fitted to a single exponential decay to obtain the rate constants summarised in table 4.10. A key describing each plot is provided within the graph.

	Exonucleolyis rates (min^{-1}) of Pfu-Pol variants			
DNA substrate	**Wild type**	**M247A**	**Y261A**	**M247A/Y261A**
ss 2APA	15.6 ± 1.9	16.1 ± 1.5	1.9 ± 0.1	1.5 ± 0.2
^{2A}PA control	3.1 ± 0.1	3.0 ± 0.3	1.9 ± 0.3	2.4 ± 0.2
^{2A}PA U+2	12.8 ± 2.3	13.7 ± 3.8	11.9 ± 2.0	15.1 ± 3.2
^{2A}PA U+4	2.4 ± 0.2	3.0 ± 0.3	1.5 ± 0.1	1.9 ± 0.3

Table 4.10: Rates of 3′-5′ exonucleolytic activity exhibited by Pfu-Pol B (wild type and mutants) acting on the DNA substrates described in table 4.1. The mean rates are provided for each condition ± the standard deviation from at least three observations.

In agreement with previously published data (Russell *et al.*, 2009), a significant increase in 3'-5' exonucleolytic activity was observed with the ^{2A}PA U+2 primer-template with each of the tested enzymes, as compared to the control. However, in contrast to previous results, the increase above the control rate was roughly 5-fold. In

general, the M247A and Y261A mutations do not appear to significantly influence the rates of 3'-5' exonucleolysis with any of the primer-templates. The only substrate to show a substantial change in k_{exo} was the single-stranded 2APA substrate, which produced a significantly reduced rate of exonucleolysis with the Y261A single and double mutants.

4.7: Discussion

Both the steady state and time-resolved fluorescence analyses clearly show an increase in 2AP fluorescence upon binding by Pfu-Pol B with both U+2 and U+4 primer-templates. However, a much more profound fluorescence enhancement is seen with U+2 than with U+4. These data are consistent with diminished stacking of the modified base as it moves from a double- to single-stranded environment. The fluorescence decay parameters provide convincing evidence for a decrease in the amount of double-stranded DNA upon binding by the enzyme, manifested by weakening of interbase stacking (and hence an increase in the lifetime of the shortest component, τ_1), as well as a large transfer of population to poorly stacked conformations. The fluorescence data found with the U+2 primer-templates, inferring strand separation, agree well with an X-ray crystal structure of Tgo-Pol B bound to a hypoxanthine +2 DNA substrate, which has since been produced (figure 4.19) (Killelea *et al.*, 2010). In contrast, a much smaller increase in 2AP fluorescence intensity and only a slight increase in the lifetime of the shortest lifetime component, τ_1, are seen with the U+4 primer-template, with the persistence of a high population of strongly stacked states. These data suggest less profound DNA distortion. Again these findings agree well with available structural data (Firbank *et al.*, 2008). It is concluded that the presence of uracil at +2 leads to more DNA unwinding by archaeal Pol B than observed with U+4, however, even with uracil located at the +4 position some partitioning may occur.

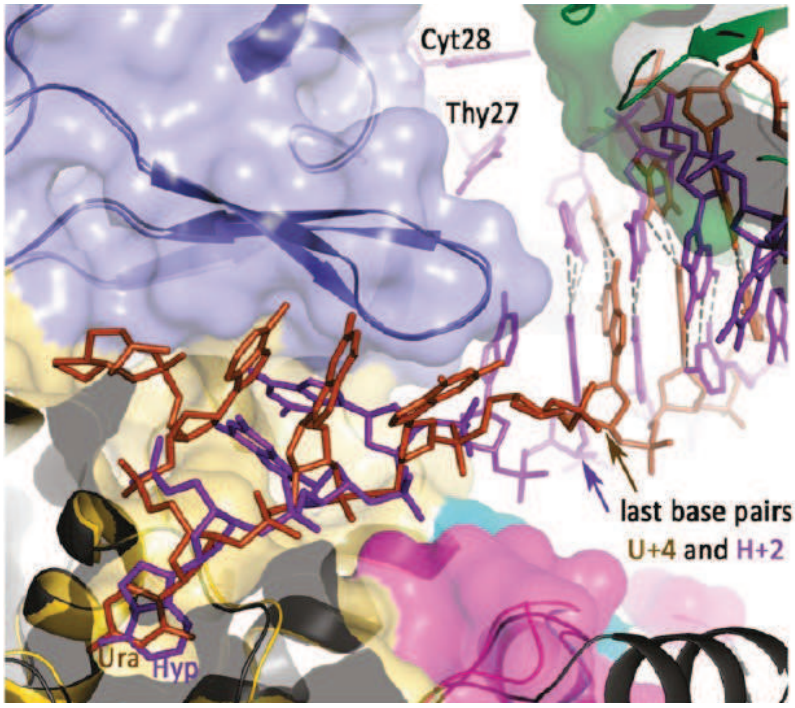


Figure 4.19: A close up image of a crystal structure of the family B DNA polymerase of *Thermococcus gorgonarius* bound to two different DNA hairpin structures. Both of the substrates were designed to simulate primer-templates, with one containing the deaminated base, uracil, at the +4 position (shown in orange) and the other containing the deaminated base, hypoxanthine, at the +2 position (shown in purple) (Firbank *et al.*, 2008; Killelea *et al.*, 2010).

Many steady state and time resolved experiments have been carried out with T4 and RB69 polymerases using 2AP-containing DNA (Subuddhi *et al.*, 2008; Bloom *et al.*, 1994; Beechem *et al.*, 1998; Hariharan and Reha-Krantz, 2005), however, exact comparisons with Pfu-Pol B are not straightforward. Both of the aforementioned polymerases belong to the B family and have obvious sequence and structural homology with Pfu-Pol B, however, the viral enzymes do not respond to the presence of uracil in template strands (Fogg *et al.*, 2002). Studies with the viral polymerases used primers with 2AP located at the extreme 3' end, as opposed to the penultimate position used in most of the analyses here. In contrast to the studies of viral polymerases, many of the fluorescence analyses were performed using a D215A (exo-) version of Pfu-Pol B, due to hydrolysis by the wild type enzyme under certain conditions. With T4 and RB69 polymerases a less pronounced increase in 2AP fluorescence is seen with the exo- enzyme, compared to the exo+ (Subuddhi *et al.*,

2008; Bloom *et al.*, 1994). However, based on the steady-state analyses, very little difference is seen between the two enzymes in this study. Despite these differences, numerous comparisons between this and previous studies are apparent. Binding by T4 and RB69 polymerases generally results in an increase in steady-state fluorescence with 2AP-containing primer-templates. Time-resolved experiments similarly demonstrate an increase in the population of components with longer lifetimes, at the expense of shorter ones (Subuddhi *et al.*, 2008; Bloom *et al.*, 1994; Beechem *et al.*, 1998; Hariharan and Reha-Krantz, 2005). Both sets of results tend to be interpreted as arising from increased primer-template strand separation on formation of 3'-5' exonuclease-competent complexes. Clearly the conclusion drawn with the viral polymerases agrees with our proposition that, following binding by Pfu-Pol B, U+2 primer-templates undergo greater strand separation than U+4. Single turnover kinetics, performed in parallel with fluorescence experiments, showed the 2AP-containing U+2 primer-template was more susceptible to proofreading exonuclease activity than the U+4 substrate.

DNA unwinding is a pre-requisite for proofreading exonucleolysis (Joyce, 1989; Reha-Krantz, 2010), indicating that the degree of strand separation seen with U+2 and U+4 correlates with the rate at which they are subject to exonucleolysis (figure 4.18, table 4.10). As discussed in the introduction and in a previous publication (Russell *et al.*, 2009), stimulation of proofreading occurs as the polymerase approaches nearer than four bases to uracil or hypoxanthine and serves as an additional safeguard to prevent copying of the pro-mutagenic base. The polymerase undergoes futile cycles of dNTP incorporation and dNMP excision, termed idling (Khare and Eckert, 2001; Khare and Eckert, 2002), without progressing beyond the deaminated base. Proofreading is normally observed following incorporation of an incorrect base during replication. The resulting mismatched base pair is unstable and susceptible to unwinding. The unstable pairing therefore enables the initiation of strand separation that necessarily precedes exonucleolysis (Reha-Krantz, 2010). During uracil recognition by archaeal polymerases none of the base pairs in the primer-template are mismatches and, as such, a fully Watson-Crick base-paired duplex must be unravelled to initiate proofreading. To elucidate features contributing to the unwinding of a stable duplex, attention has focussed on two amino acids suggested to be important based on structural studies (Firbank *et al.*, 2008; Killelea *et al.*, 2010). The amino

acid at position 247 (arginine in Tgo-Pol B or methionine in Pfu) is at the tip of the β-hairpin, a structural element implicated in primer-template strand separation (Shamoo and Steitz, 1999; Reha-Krantz, 2010; Trzenecka *et al.*, 2009; Subuddhi *et al.*, 2008; Hogg *et al.*, 2007). As shown in figure 4.20, R247 stacks against one of the newly formed single-stranded template bases that arise following partial primer-template denaturation. It was earlier proposed that this amino acid may act as a wedge to pry apart the strands (Killelea *et al.*, 2010). A second amino acid, tyrosine 261, is located on an α-helix and positioned between separated primer and template bases (figure 4.20), possibly preventing their re-annealing (Killelea *et al.*, 2010).

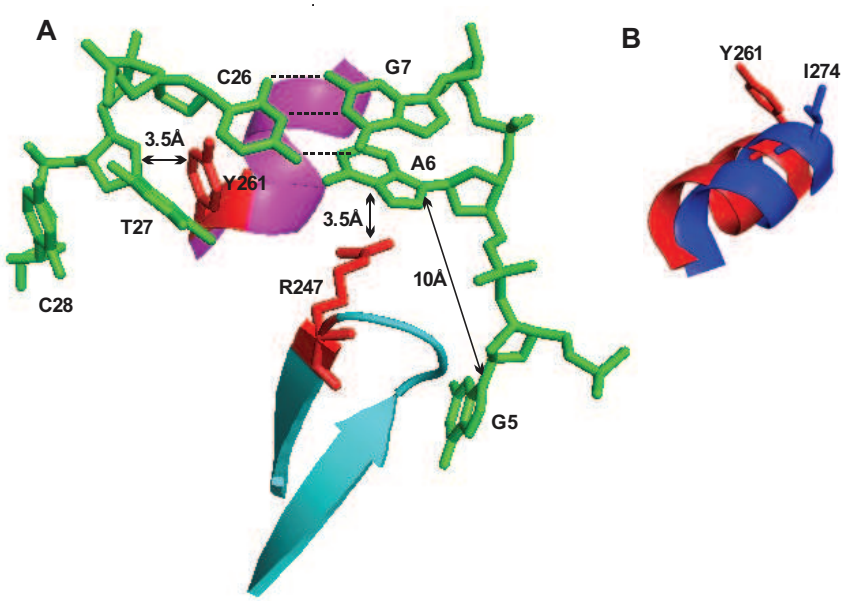


Figure 4.20: A: Interaction of Tgo-Pol B Y261 and R247 with a bound DNA structure (Killelea *et al.*, 2010). C26 and G7 are base paired but the primer bases T27 and C28 are single-stranded and unwound from their complementary template bases, A6 and G5, respectively. Y261 (red) is located on an α-helix (magenta) and is situated ~3.5 Å from T27. R247 (red) is located on the β-hairpin motif (cyan) ~3.5 Å from A6. The N9 atoms of the neighbouring bases, A6 and G5, are about 10 Å apart. **B:** Superimposition of Tgo-Pol B Y261 (red) with RB69 I274 (blue). Both amino acids are located on the same face of an α-helix. Both structures are taken from editing complexes: Tgo-Pol B (PDB ID, 2xhb) (Killelea *et al.*, 2010); RB69-Pol (PDB ID 1clq) (Shamoo and Steitz, 1999).

Both M247 and Y261 have been converted to alanine on their own and in combination with one another, in both a wild type and exo- (D215A) background of Pfu-Pol B. In general, following binding of both U+2 and U+4 primer-templates, the M247A, Y261A and M247A/Y261A mutant enzymes exhibited similar behaviour to the wild type enzyme, in terms of strand separation. However, in every case binding by the mutants resulted in a markedly smaller increase in steady-state 2AP fluorescence, compared to the wild type control (figures 4.14-4.17). The three mutants did not decrease proofreading rates seen with primer-templates, as assessed using single turnover kinetics (table 4.10). M247A additionally did not alter the hydrolysis rates seen with single-stranded DNA. However, a ~10-fold decrease in the rate of degradation of this substrate was seen with Y261A and the double mutant. The fluorescence and kinetic results seen with the mutants complicate the simple conclusion arrived at for the wild type enzyme with U+2 and U+4 substrates; namely that increased strand separation straightforwardly gives rise to faster proofreading. The reduced increase in 2AP fluorescence seen with M247A and Y261A suggests that more of the primer-template remains double-stranded following binding of the mutants, compared to the wild type enzyme. Therefore, as suggested by structural studies, both amino acids appear to play a role in DNA denaturation. However, the diminished amount of separated primer-template is not reflected in a drop in the rate of proofreading exonucleolysis. The data in table 4.10 were determined under single turnover conditions, which measure the slowest step in the reaction pathway following DNA binding up to, and including, bond cleavage. Strand separation has been suggested to be the rate-limiting step for proofreading exonucleolysis (Reha-Krantz, 2010), compatible with the basic idea that rapid exonucleolysis, as seen when the wild type enzyme acts on the U+2 substrate, directly arises from enhanced primer-template denaturation. However, under such a scenario, mutations that reduce primer-template unwinding, such as M247A and Y261A, would also be expected to lower rates of exonucleolysis. While, at present, a full rationalisation of these results is not possible, a similar combination has been seen with RB69 polymerase and β-hairpin-located amino acids. Removal of the entire loop region of the β-hairpin resulted in diminished enhancement of 2AP fluorescence (Subuddhi *et al.*, 2008), yet did not lead to a reduction in the rate of proofreading exonucleolysis (Hogg *et al.*, 2007). Rather, for each individual DNA binding event, the deleted loop mutant hydrolysed a lower percentage of the primer-template than the wild type enzyme. It was proposed that

the role of the β-hairpin is to tightly bind DNA during proofreading and that its deletion increases the probability of dissociation, thus reducing product yield. With the Pfu-Pol B mutants a drop in the amount of proofread product was not observed. This may arise as experiments were carried out in the presence of PCNA, which reduces primer-template dissociation. Uracil-containing DNA also possesses an intrinsically high affinity for archaeal family B polymerases (Shuttleworth *et al.*, 2004; Gill *et al.*, 2007).

Further information about the significance of Pfu-Pol B M247 and Y261 is available from sequence comparisons. The amino acid in the β-hairpin at position 247 does not show especially pronounced conservation (figure 4.21). Even when amino acid similarity searches are confined to the two closely related genera, Pyrococcus and Thermococcus, a degree of variation is observed, with arginine in 21 species, methionine in 8 and serine in 9 (Killelea *et al.*, 2010). When the wider phylum euryarchaeota is considered, no retention of the amino acid at this position can be discerned. With Tgo-Pol B, R247 inserts between two neighbouring template bases and may either force them apart or, alternatively, stabilise the distorted structure (figures 4.12 and 4.20). While the similarly sized methionine found in Pfu-Pol B could conceivably play the same role, it is difficult to see how the much smaller serine, seen with several Pyrococcus/Thermococcus polymerases, could fulfil such a function, unless compensatory changes are seen elsewhere in the motif. Site directed mutagenesis often involves conversion to alanine, however, the side chain of this amino acid is similar to that of the naturally-occurring serine residue, thus the polymerase variant M247A may not be strongly disabled. With the RB69 polymerase, R260 interacts with both primer and template bases and occupies a location that approximates to that seen with the archaeal amino acid 247 (Shamoo and Steitz, 1999). At present it is unclear if R260 is a strict homologue of residue 247; their interaction with DNA is different (Killelea *et al.*, 2010; Shamoo and Steitz, 1999) and spatial superimposition of both the amino acid itself and the entire β-hairpin is not exact. Furthermore, it has been observed that the β-hairpin tends to show conservation at the secondary, rather than primary, structural level (Reha-Krantz, 2010; Trzenecka *et al.*, 2009) and therefore changes to individual amino acids within this motif may have only marginal affects. A much higher level of conservation is seen with the α-helix-located amino acid, Y261. Considering the entire euryarchaeal phylum, this residue is

invariably aromatic or large and aliphatic: 51 species with tyrosine, 12 with phenylalanine, 12 with tryptophan and 21 with leucine (Killelea *et al.*, 2010). All of these amino acids possess bulky hydrophobic side chains. The proposed role of amino acid 261 in keeping apart separated primer-template strands could clearly be fulfilled by all of these large side chains. With RB69 polymerase I274 may play the role of residue 261 in archaeal Pol B, as structural comparison shows pronounced spatial overlap of these residues, as well as the α-helix on which they lie (figure 4.17). In viral polymerases, like the archaea, this amino acid is overwhelmingly aromatic or possesses a large hydrophobic side chain. Three editing-mode RB69:DNA structures have been described and show that I274 is near the separation point of the primer and template strands (Shamoo and Steitz, 1999; Hogg *et al.*, 2004). However, this amino acid does not appear to directly contact the DNA or aid strand separation and seems to play a more passive role than that described for Y261 in Tgo-Pol B and Pfu-Pol B.

Pfu:	231	TIGRDGSEPKMQRIGDMTAVEVKGRIHFDLYHVITRT--INLPTYTLEAVYE	280
Tch:	111	PLGRDGSEPKMQRLGDMTAVEIKGRIHFDLYHVIRRT--INLPTYTLEAVYE	159
Pgl:	232	PLGRDGSEPKMQRLGDMTAVEIKGRIHFDLYHVIRRT--INLPTYTLEAVYE	280
Tgo:	231	ILGREGSEPKIQRMGDRFAVEVKGRIHFDLYPVIRRT--INLPTYTLEAVYE	280
Tgu:	231	TLGRDGSEPKIQRMGDRFAVEVKGRIHFDLYPVIRRT--INLPTYTLEAVYE	280
Tga:	231	TLGRDGSEPKIQRMGDRFAVEVKGRIHFDLYPVIRRT--INLPTYTLEAVYE	280
Tko:	231	ALGRDGSEPKIQRMGDRFAVEVKGRIHFDLYPVIRRT--INLPTYTLEAVYE	280
Pya:	232	PIGRDGSEPKMQRMGDGFAVEVKGRIHFDIYPVIRRT--INLPTYTLEAVYE	280
T4:	241	PIGRVKSKLIQNMYGSKEIYSIDGVSILDYLDLYKKFAFTNLPSFSLESVAQ	292
RB69:	247	PHRKTRVKVIENMYGSREIITLFGISVLDYIDLYKKFSFTNQPSYSLDYISE	295

Figure 4.21: Amino acid similarity within the ß-hairpin region across a number of family B DNA polymerases belonging to members of the *Thermococcales* order, as well as the viral family B enzymes of bacteriophages T4 and RB69. Residues shown in red show a high degree of conservation. Residues shown in blue show a moderate degree of conservation. The two mutated amino acids in Pfu-Pol B, M247 and Y261, and their equivalent residues are highlighted in yellow. The full names and abbreviations of each of the organisms are provided in the abbreviations section on pages xix and xx.

In summary 2AP fluorescence has been used to demonstrate that archaeal family B DNA polymerases denature primer-templates as they approach uracil and, at least on

2AP evidence, two amino acids (methionine 247 and tyrosine 261 in Pfu-Pol B) have been identified that contribute to unwinding. The capture of uracil or hypoxanthine by the deaminated-base binding pocket is almost certainly the source of binding energy used for unwinding the fully complementary primer-template (Russell *et al.*, 2009). However, the pathway that commences with the encounter of uracil at +2, followed by proofreading exonucleolysis and re-set of uracil to the +4 position, as well as the roles of individual structural elements in the process, remain obscure; particularly the relationship between primer-template unwinding and proofreading exonuclease activity. Overall, the enzyme performs a remarkable job to proofread, rather than extend, a fully Watson-Crick base-paired primer-template, thereby preventing copying of deaminated bases in the template strand. Currently attempts are being made to study the mechanism more fully. It is suggested that two approaches are likely to be fruitful. Firstly, stopped flow fluorescence with the 2AP primer-templates should lead to elucidation of individual steps in the overall reaction scheme. Secondly, the isolation of mutants that are still able to tightly bind uracil, but no longer show differences in strand separation and hydrolysis between U+4 and U+2, may further define the role of individual amino acid residues.

Chapter 5

In vivo analysis of deaminated base recognition and 3'-5' exonuclease activity by Pol B in *Thermococcus kodakarensis*

5.1: Background

All of the work presented in chapter 5 is the result of a collaboration with the Reeve/Santangelo group based at Ohio State University, Columbus, USA.

To complement the biochemical studies on deaminated base recognition and exonuclease activity presented in chapters 3 and 4, a genetic approach has been undertaken to enable investigation of the *in vivo* effects of these activities in an archaeon. Until fairly recently a dearth of systems and techniques existed for genetic investigation of the archaea (Allers and Mevarech, 2005). Within the last ten years, however, it has been discovered that species such as *Thermococcus kodakarensis* and *Sulfolobus islandicus* are, in fact, highly amenable to genetic manipulation (Santangelo *et al.*, 2010; She *et al.*, 2009). The focus of this chapter will therefore be the model organism, *T. kodakarensis*, which belongs to the Euryarchaeota subdomain.

The genetic utility of *T. kodakarensis* stems from its natural competency for DNA uptake and its ability to efficiently incorporate donor DNAs into its chromosome (Sato *et al.*, 2003). Short regions of homology between the chromosome and donor DNA are sufficient to drive integration by homologous recombination. Critically, transformation does not require the employment of protoplasts or exacting experimental techniques. As such, *T. kodakarensis* represents an ideal candidate for studies involving complex genetic modifications.

As with most members of the Euryarchaeota, *T. kodakarensis* is thought to possess two functional, replicative DNA polymerases: a monomeric family B DNA polymerase and a heterodimeric family D DNA polymerase (Henneke *et al.*, 2005). The roles of the two polymerases have not been conclusively shown, however, it is known from targeted gene deletions that both are essential in a halophilic euryarchaeon (Berquist *et al.*, 2009). Based on their *in vitro* strand displacement activities, it has been suggested that euryarchaeal Pol B and Pol D may be responsible for leading and lagging strand DNA replication, respectively (Henneke *et al.*, 2005). This model would be highly analogous to the mode of replication employed by eukaryotes, in which Pol ε and Pol δ have been shown to perform the corresponding

roles, following initiation by Pol α (Garg and Burgers, 2005; Nick McElhinny *et al.*, 2008; Pursell *et al.*, 2007). At present, however, conclusive genetic evidence does not exist to demonstrate the roles of Pol B and Pol D in a euryarchaeon.

If, as is assumed to be the case, Pol B *is* responsible for replicating ~50 % of the euryarchaeal genome, then manipulation of the Pol B-encoding gene to produce a mutagenic phenotype should result in a significant increase in the accumulation of chromosomal mutations. The work presented in chapter 5 describes the production of stable mutant strains of *T. kodakarensis* by double homologous recombination. The construction of vectors, containing the Pol B-encoding TK0001 gene of *T. kodakarensis*, is described, followed by the introduction of specific mutations to TK0001. The resulting plasmids were transformed into *T. kodakarensis* to produce a variety of stable mutant strains containing altered copies of TK0001. On the basis of previous studies, a V93Q mutation was introduced to TK0001 to enable the effects of deaminated base recognition by Pol B to be investigated. The V93Q mutation has previously been shown *in vitro* to completely eliminate the specific interaction of Pfu-Pol B with uracil and hypoxanthine (Fogg *et al.*, 2002). Critically, the V93Q variant of Pfu-Pol B retains both polymerising and 3'-5' exonucleolytic activities. By introducing this mutation to TK0001 and employing the resulting strain in a reporter assay, it has been possible for the first time to study the effects of deaminated base recognition by archaeal Pol B *in vivo*.

The study was further extended to investigate the influence of the 3'-5' exonucleolytic activity of Pol B by introducing a D215A mutation to TK0001 (Evans *et al.*, 2000). This mutation has previously been shown *in vitro* to almost completely eliminate the proofreading exonucleolytic activity of Pfu-Pol B. The mutation was therefore introduced on its own and in combination with V93Q to produce 4 stable mutant strains. The plasmids and strains are described in greater detail in chapter 5.2.

5.2: Construct preparation and production of mutant *T. kodakarensis* strains TR1-TR4

In order to analyse the *in vivo* effects of deaminated base recognition and 3'-5' exonucleolytic activity by Pol B in a euryarchaeon, stable mutant strains of *T. kodakarensis* were produced based on *T. kod* KW128 (Sato *et al.*, 2004). KW128 is a tryptophan auxotrophic mutant, so by promoting allelic exchange with an artificial vector containing the TrpE gene, then growing the transformed cells on minimal plates lacking tryptophan, it is possible to select for positive transformants that contain donor DNA. Plasmids were constructed based on pUMT2 (Sato *et al.*, 2004) to contain the TK0001 gene, which encodes Pol B, together with the upstream genomic architecture of TK0001; in this case TK2303, TK2304 and TK2305. The two pieces of DNA were cloned either side of the TrpE selectable marker and the resulting vector renamed pTR1 (figure 5.1). D215A and V93Q mutations were then introduced to the TK0001 gene, both on their own and in combination with one another, to eliminate 3'-5' exonucleolytic activity and/or deaminated base recognition, respectively (Evans *et al.*, 2000; Fogg *et al.*, 2002). The vector containing the D215A mutation was renamed pTR2. The vector containing the V93Q mutation was renamed pTR3. The vector containing both mutations was renamed pTR4.

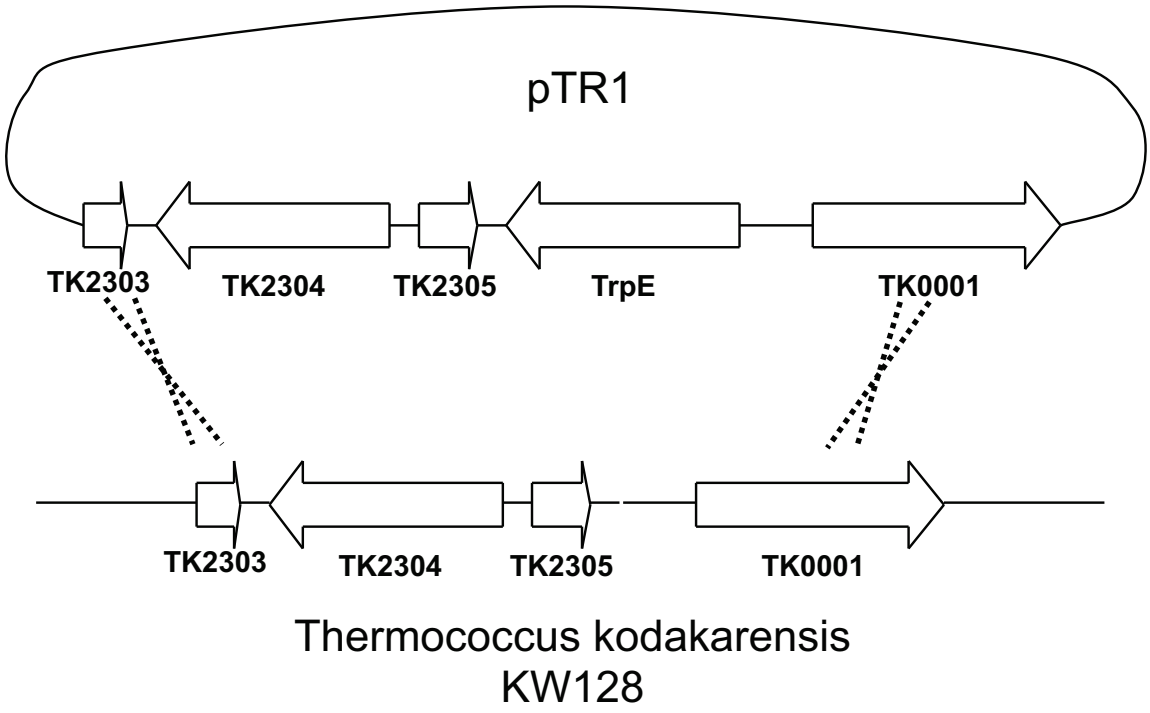


Figure 5.1: Transformation of *T. kodakarensis* KW128 using pTR1 to promote double homologous recombination.

Once the four plasmids had been constructed and confirmed by full sequencing, they were used to perform transformations of *T. kodakarensis* KW128. Since the likelihood of introducing a given mutation decreases with distance from the marker, a greater number of pTR2 and pTR4 transformants were grown and sequenced. After confirming that the desired mutation(s) had been introduced by double homologous recombination (i.e. that one recombination event occurred at two sites, either side of the TrpE marker), diagnostic PCRs were performed to confirm the absence of wild type KW128 DNA. The confirmed strains were re-named TR1-TR4, corresponding to their respective plasmids, and are summarised in table 5.1. TR1-TR4 were then used in the assays described in chapters 5.3 and 5.4.

Name of strain	TK0001 genotype	Predicted phenotype
TR1	Wild type	Wild type
TR2	D215A	3'-5' exo-
TR3	V93Q	Uracil-insensitive
TR4	D215A/V93Q	3'-5' exo-/uracil-insensitive

Table 5.1: Summaries of the four mutant strains of *T. kodakarensis* used in this study. The name of each strain is given in column 1, the TK0001 genotype of each strain is indicated in column 2 and the phenotype of the family B DNA polymerase of each strain, as predicted by *in vitro* analysis, is indicated in column 3.

5.3: Growth curves of mutant and wild type strains

The data presented in chapter 5.3 were produced by Dr Tom Santangelo and colleagues.

In order to confirm that the newly constructed strains, TR1-TR4, were healthy and viable and behaved in a manner comparable to wild type *T. kodakarensis*, growth curves were produced by measuring the optical density at 600 nm (OD_{600}) of actively growing cultures over an 11-hour time course. Reference to figure 5.3 reveals that all four strains exhibited exponential growth, reaching saturation after approximately 8 hours. The growth curves confirm that the four strains behaved more or less identically to one another.

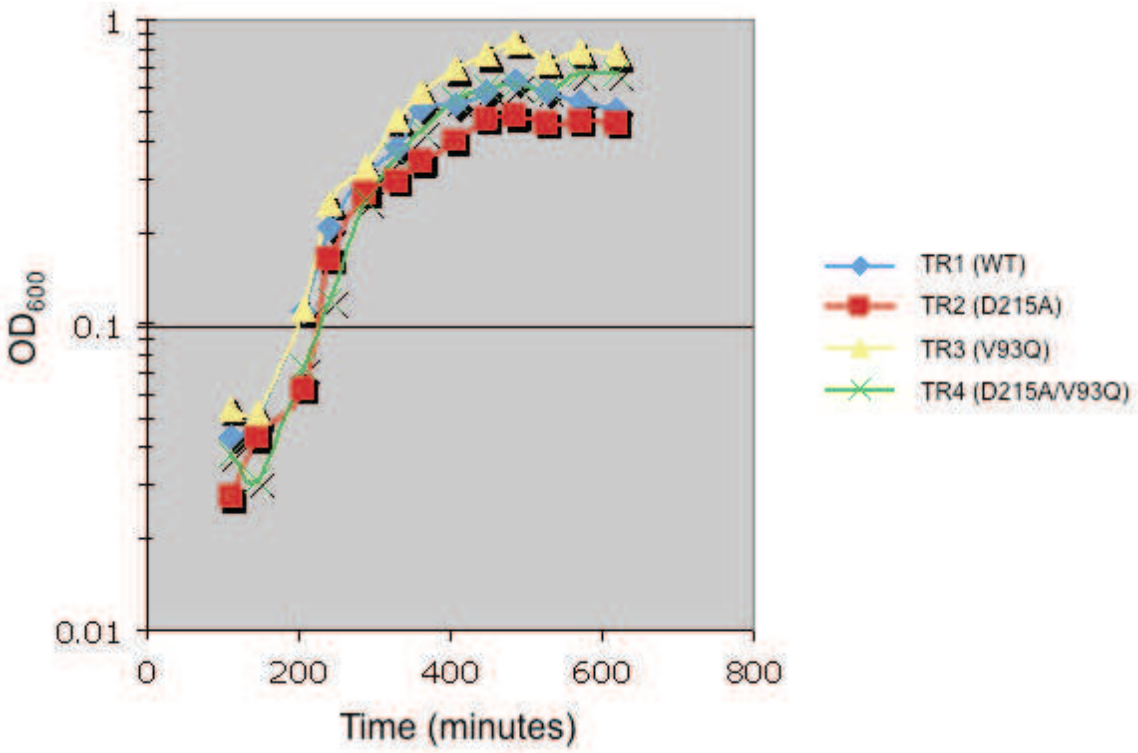


Figure 5.2: Growth curves for mutant *T. kodakarensis* strains, TR1-TR4, cultured at 85 °C over an 11 hour time course. The optical density of the four cultures was measured at 600 nm and plotted on a logarithmic scale against the elapsed time post inoculation.

5.4: Application of a novel reporter assay for the production of mutational profiles for strains TR1-TR4

In order to establish what types of mutations were produced by each of the four mutant strains, a novel reporter assay was employed. Traditionally, the most common reporter assays are based on the bacterial *lacZ* gene, providing a simple colourimetric indication of the status of the gene (Miller, 1972). In the standard *lacZ* assay, disruption of the reporter gene prevents expression of β-galactosidase, which is capable of metabolising X-gal, a colourless sugar, to 5-bromo-4-chloro-indoxyl. Under aerobic conditions 5-bromo-4-chloro-indoxyl is spontaneously oxidised to form the bright blue pigment, 5,5'-dibromo-4,4'-dichloro-indigo, thereby providing a

clear visual indication of expression of the gene. Unfortunately, due to the requirement for oxygen and the lack of a β-galactosidase gene in *T. kodakarensis*, the traditional *lacZ* assay is not easily applicable. Any *lacZ* reporter assay of DNA replication would require the reporter gene to be housed on a shuttle vector. The vector would subsequently require shuttling through both *T. kodakarensis* and *E. coli*, thereby complicating any possible conclusions. Another commonly used reporter gene is that of the green fluorescent protein (GFP) from *Aequorea victoria* (Jacobi *et al.*, 1998). Unfortunately, since the protein is derived from a mesophilic organism and also requires oxygen, it is not ideally suited to *in vivo* studies of *T. kodakarensis*. Recent advances have, however, been made in the production of thermostable derivatives (Endoh *et al.*, 2008). Unfortunately, as such, there is currently no obvious colourimetric reporter gene available for studies of thermophilic anaerobes that lack a β-galactosidase gene. A novel reporter assay, based on the TK0664 gene, was therefore employed in this study to enable selection of knockout mutations resulting from chromosomal replication errors.

The TK0664 gene encodes hypoxanthine/guanine phosphoribosyltransferase (HGPRT) and enables *T. kodakarensis* to scavenge purines when grown on minimal media. Ordinarily, HGPRT converts guanine (or hypoxanthine) to GMP (or IMP), as shown in figure 5.3 (Berg *et al.*, 2007). GMP may subsequently be converted to GTP and later incorporated into DNA. Disruption of TK0664 prevents purine scavenging due to perturbation of the metabolic pathway. The purine analogue, 6-methylpurine (6MP), can similarly be acted upon by HGPRT, however, incorporation of 6MP derivatives into DNA is severely cytotoxic (Parker *et al.*, 1998). By growing *T. kodakarensis* in minimal media and in the presence of 6MP, it is therefore possible to screen for knockout mutations in TK0664, since knockout mutants are no longer susceptible to the cytotoxic effects of scavenged 6MP (Santangelo *et al.*, 2010). After screening for knockout mutations, individual mutants were isolated, grown in liquid culture and subjected to genomic DNA extraction. The purified DNA was then used to amplify TK0664, including an additional 100 bp upstream and downstream of the gene. The resulting DNA was then sequenced to produce mutational profiles for each of the newly constructed strains.

Guanine + phosphoribosyl pyrophosphate (PRPP) → guanylate (GMP) + pyrophosphate (PP_i)

Hypoxanthine + PRPP → inosinate (IMP) + PP_i

Figure 5.3: Metabolism of guanine and hypoxanthine to GMP or IMP, catalysed by the action of hypoxanthine-guanine phosphoribosyltransferase (Berg *et al.*, 2007).

5.4.1: Mutation rates of TR1-TR4

Attempts were made to measure the mutation rates for each of the four mutant strains. The TK0664 reporter assay previously described was employed to establish the number of TK0664 knockout mutants in a given culture. The same culture was simultaneously used to perform serial dilutions on a minimal Trp- plate (lacking 6MP) to produce an accurate count of the number of viable cells. The intention was to compare the number of knockout mutations with the total number of viable cells in order to calculate the approximate mutation rate of each strain. Unfortunately, despite the relative simplicity of the technique, producing accurate cell counts proved unsuccessful as no viable cells were ever observed from the serial dilutions. The only possible explanation that we are able to offer is that the osmotic potential of the artificial seawater, which was used to perform the dilutions, may have been too high and caused the cells to rupture. An approximation of the number of cells was therefore made, based on the OD_{600} of grown cultures. From those estimates, growth on 6MP-containing plates led to a reduction in the number of viable cells by a factor of roughly ~10^6-10^8. Attempts to accurately determine the mutation rates of the four strains were therefore unsuccessful. The focus of the investigation was instead placed on determining what types of mutation occurred most frequently in each of the strains.

5.4.2: The "jackpot" effect

Initially, 8-10 mutant colonies were picked from each 6MP plate and their TK0664 genes subsequently sequenced, however, after analysing the first few dozen mutants it became apparent that the same mutations were occurring multiple times on the same plate. In extreme cases every TK0664 mutant from a given plate possessed the same mutation. Rather than occurring independently of one another, it was assumed that the mutants were sisters/daughters of one another and that the mutations had occurred early on during culturing, proliferating exponentially. In contrast, when mutations arose later on during culturing a more even distribution of mutations was seen. The observed phenomenon was therefore termed the "jackpot" effect. In order to counter this, it was decided that only 1-5 mutant colonies would be picked from each plate and that the rest would be discarded. Colonies derived from the same culture and exhibiting the same genotype were assumed to be daughter cells derived from the same mutant and were counted only once.

After analysing the first 3 plates of TK0664 mutants from the TR2 (exo-) strain, it was observed that one mutation, a g>a transition 9 bases upstream from the start codon, was present in over 90 % of sequences, regardless of which starter culture the mutants were derived from. It was assumed that the mutation occurred early on during amplification of the parental culture shortly after transformation of the KW128 *T. kodakarensis* and that the repeats were in fact daughters of the same mutant. To prevent the g>a transition occurring in subsequent plates, it was therefore necessary to "clean-up" the parental stock of TR2. In order to counter the "jackpot effect" and ensure that every starter culture was monoclonal and derived from a wild type TK0664 colony, serial dilutions of TR2 and the other three strains were performed on minimal Trp- plates lacking 6MP to produce single colonies. 5-10 single colonies from each strain were then picked and resuspended in artificial seawater, with half of the resuspension plated on Trp- plates and half on Trp-/6MP plates. After three days' growth the two sets of plates were compared and, provided there were no colonies on the corresponding Trp-/6MP plate, it was assumed that the original colony was wild type for TK0664. Single colonies were then picked from the corresponding Trp- plate (without 6MP) and used to inoculate fresh starter cultures. The "cleaned-up" starter

cultures were therefore all monoclonal and possessed wild type copies of TK0664. "Clean" cultures were then grown, plated on 6MP-containing plates to select for TK0664 knockout mutations and the mutations subsequently analysed, as previously described. When mutations were seen at the same locus, but arose in different cultures, it was assumed that they had occurred independently of one another and were the result of mutational hot spots, rather than being sisters or daughters of one another.

5.4.3: Mutational profiles of TR1-TR4

At biologically relevant temperatures spontaneous deamination of DNA is predicted to affect cytosine residues 50-500 times more frequently than the other affected canonical bases (Schroeder and Wolfenden, 2007). Spontaneous deamination of cytosine results in the production of uracil, which functions as an efficient template for adenine incorporation by most DNA polymerases (Takasawa *et al.*, 2004; Vaisman *et al.*, 2006). Therefore, in order to gain an indication of the effects of spontaneous DNA deamination on the four newly-constructed strains, particular attention was paid to the frequency of c>t, or conversely g>a, transition mutations, depending on which strand replicative bypass of uracil was assumed to have occurred in. The summarised data from the TK0664 reporter assay, including the c>t/g>a transition rate for each strain, can be seen in table 5.2. The exact mutations seen for each strain are described in appendix 2 and appendix 3.

Strain	Sample size	c>t/g>a transitions	t>c/a>g transitions	Transversions	deletions	insertions	% c>t/g>a transitions
TR1 (WT)	102	21	11	9	38	20	21
TR2 (exo-)	91	23	10	9	42	7	25
TR3 (uracil-insensitive)	66	18	4	12	24	6	28
TR4 (double mutant)	64	18	1	16	25	3	29

Table 5.2: Frequency of mutations produced by four mutant strains of *Thermococcus kodakarensis*. The strains were all based on *T. kodakarensis* KW128 and contain modifications to the TK0001 gene. Full descriptions of each of the strains are provided in chapter 5.2.

From the data presented in appendices 2 and 3, the c>t/g>a transition rate for each strain appears to be highly variable between batches of mutants that were isolated on different days. The c>t/g>a rate for the TR1 (wild type) strain, for example, was 9 % (n= 35) for the samples that were analysed in 2010. However, the same strain analysed in 2011 produced a c>t/g>a transition rate of 28 % (n = 64). Similar levels of batch-to-batch variation were recorded for all of the strains. In spite of this variation, all of the strains with pro-mutagenic copies of Pol B (TR2-TR4) show elevated rates of c>t/g>a transition mutations. The c>t/g>a transition rate for the TR2 (exo-) strain was 25 %, compared to 21 % for the TR1 (wild type) strain. The c>t/g>a transition rate for the TR3 (uracil-insensitive) strain was higher still, at 28 %, and the TR4 double mutant strain exhibited the highest rate of all at 29 %. It was decided that, although the 7 % increase observed between the TR1 (wild type) and TR3 (uracil-insensitive) strains was notable, it was not worth performing further statistical analyses on the data due to the significant variation previously described. Despite doubts about the quality of the data, a number of other notable observations can be made. The frequency of t>c/a>g transition mutations, as would occur when replicative bypass of deaminated adenine had taken place, did not correlate with deaminated base recognition and was significantly higher in the TR1 (wild type) and TR2 (exo-) strains.

In order to determine whether or not there was a strand-specific bias in c>t/g>a transition mutations, the numbers of c>t and g>a transition mutations were compared

in the coding and non-coding strands of TK0664 (table 5.3). Interestingly, an extreme bias was observed for all four strains with the cytosine residue located in the non-coding strand in 99 % of cases (n = 80). Only one transition occurred that would indicate uracil bypass in the coding strand and was observed in the TR4 strain.

strain	sample size	c>t/g>a transitions	% of c>t/g>a transitions with c in non-coding strand	% of c>t/g>a transitions with c in coding strand
TR1 (WT)	102	21	100	0
TR2 (exo-)	91	23	100	0
TR3 (uracil-insensitive)	67	18	100	0
TR4 (double mutant)	64	18	94	6

Table 5.3: Frequency of cytosine>adenine transition mutations in the coding and non-coding strands of the TK0664 reporter gene for four mutant strains of *Thermococcus kodakarensis*. The strains are fully described in chapter 5.2.

5.5: Discussion

The results presented in chapter 5.4 provide some evidence for Pol B's role in replicating the chromosomal DNA of euryarchaea, however, as discussed in chapter 5.4.3, doubts remain over the quality of the data. Despite the considerable variability, the data appear to suggest that deaminated base recognition by archaeal Pol B may contribute to maintaining genomic stability through the prevention of c>t/g>a transition mutations. The uracil-insensitive TR3 strain shows an increase of 7 % in the frequency of these mutations, indicating that replicative bypass of deaminated cytosine occurred more frequently than in the wild type control. Whether or not the difference can legitimately be described as significant, however, is debatable, as statistical analyses have not been performed. Based on the data presented in chapter 5.4.3, all of the pro-mutagenic strains, TR2, TR3 and TR4, showed elevated rates of c>t/g>a transition mutations, providing some support for the theory that the combined

action of deaminated base recognition and 3'-5' exonucleolytic activity significantly reduce the rate of replicative bypass of uracil by archaeal Pol B.

As well as the aforementioned variability of the data, the observed phenotypes of the pro-mutagenic strains, in terms of the numbers of c>t and g>a transition mutations, were not nearly as dramatic as anticipated. For example, the observed rate for the V93Q/D215A double mutant was only 1 % higher than that of the uracil-insensitive single mutant. Furthermore, the rate of a>g/t>c transition mutations, as would be predicted to occur when replicative bypass of deaminated adenine takes place, did not correlate with Pol B's ability to recognise deaminated bases. This finding was less surprising, however, considering the lower frequency of spontaneous adenine deamination expected to occur at 85 °C, as compared to cytosine (Schroeder and Wolfenden, 2007). One possible explanation for the relatively small difference in c>t/g>a transition rates between the four strains is that replicative bypass of deaminated bases represents a comparatively small proportion of all mutagenic events that occur during DNA replication. As such, the relative contribution of c>t and g>a transitions to the total number of mutations may be negligable, especially compared to the increase in other types of mutation that would be expected to occur with the exo- strains, TR2 and TR4. The TR2 (exo-) strain, for example, appears to show a significantly higher rate of deletions than all of the other strains, with deletions accounting for 46 % of all of the observed mutations. TR1 shows a 37 % rate for deletions, TR3 exhibits a 36 % rate and TR4 a 39 % rate. In order to gain a greater understanding as to the relative contributions of deaminated base recognition and 3'-5' exonucleolytic activity in maintaining genomic stability, a more accurate measure of the mutation rate for each strain would need to be produced. To achieve this end a *lacZ* type assay could be employed in future studies, as was performed by Cohen-Kupiec and colleagues (Cohen-Kupiec *et al.*, 1997). In their study the anaerobically grown cells were retroactively exposed to oxygen to produce a colourimetric indication of the status of the gene. Oxygen exposure would, however, compromise the viability of the cells, rendering them useless for further investigation. An additional benefit of employing the *lacZ* gene in future studies is the fact that significant work has been undertaken to characterise the effects of every possible mutation in this gene (Bebenek and Kunkel, 1995). The TK0664 reporter assay described in this chapter selects only for knockout mutations. As such, frameshift and

nonsense mutations appear to be over-represented in the results presented in table 5.2. For example, g>a transition mutations in tyrosine result in the production of a premature stop codon and comprise a significant proportion of all of the observed mutations reported here. The only way to detect silent mutations in TK0664 is when they appear as "passenger mutations", as occurs when more than one mutation is observed in a single sequence.

The transversion rate (i.e. the frequency of purine>pyrimidine conversion, or vice-versa) for the V93Q single and double mutant strains shows the most profound change of all of the observed types of mutation compared to the wild type strain, increasing from 9 % for TR1 to 18 % for the TR3 (V93Q) strain and 25 % for the TR4 (V93Q/D215A) strain. An increase in transversion mutations could be indicative of an increase in the rate of replicative bypass of xanthine, since cytosine insertion is most commonly observed opposite xanthine (Eritja *et al.*, 1986), resulting in a g>c transversion mutation where deamination of guanine has occurred. However, due to the predicted low frequency of spontaneous guanine deamination (Schroeder and Wolfenden, 2007) and the null recognition of xanthine by archaeal Pol B in *in vitro* assays, we are currently unable to speculate as to the reasons for the observed increase in these types of mutations with the TR3 and TR4 strains.

Undoubtedly the most interesting finding of this study is the strand specific bias for c>t transition mutations in TK0664 (table 5.3). 99 % of these mutations were observed in the non-coding strand, indicating that replicative bypass of uracil occurred overwhelmingly in this strand. Since the V93Q mutant phenotype (i.e. the increase in c>t transitions) was only observable in one strand, it was initially proposed that the data provided evidence for Pol B's role in replicating just one strand of genomic DNA. If that were the case, the polymerase responsible for replicating the other strand (assumed to be Pol D) would have to be extremely efficient at recognising and responding to uracil. Whilst this may be true, the data also provide strong evidence for a transcription-coupled repair (TCR) system in the archaea. This finding would be of particular significance as an archaeal TCR system has yet to be identified (Eisen and Hanawalt, 1999). Transcription-coupled repair pathways in bacteria and eukaryotes have been known about for decades and are relatively well characterised (Selby and Sancar, 1999; Hanawalt, 1994). It has been argued, however,

that there is no evidence to support the existence of such a system in the archaea (Dorazi *et al.*, 2007; Romano *et al.*, 2007). In both of the aforementioned studies on archaeal TCR, only the effects of UV-induced DNA damage were investigated. It is possible that damage resulting from spontaneous deamination of DNA may be more pertinent to members of the archaea, given their propensity for deep-sea geothermal vents. Further support for the existence of a TCR system coupled to deaminated base recognition in the archaea is provided from alignments of the N-terminus of Pfu-Pol B (in which the uracil-binding pocket resides) with Pfu-RNA Pol. Several of the subunits of Pfu-RNA Pol show small but significant levels of homology with the N-terminus of Pfu-Pol B. It seems highly probable that archaea, like bacteria and eukaryotes, would also possess a TCR pathway, considering the common evolutionary origins of the three domains and the similar evolutionary pressures that each is likely to have faced. In order to further investigate the existence of such a pathway, biochemical analysis of Pfu-RNA Pol with DNA substrates containing deaminated bases would likely prove fruitful.

To provide stronger evidence for the roles of Pol B and Pol D in replicating chromosomal DNA and/or the presence of a TCR system in the Euryarchaeota, a number of approaches could be made. Targeted mutagenesis of the DNA polymerase active sites *in vivo* to produce stronger mutagenic phenotypes for the two enzymes, based on the results of *in vitro* assays, would be useful to enable the observation of a strand-specific bias in reporter assays. Switching the location and orientation of the reporter gene itself and observing the frequency of c>t transition mutations in each strand is also likely to an effective means of determining the role of Pol B in chromosomal DNA replication and establishing whether or not a TCR system exists in euryarchaea. If the overwhelming majority of c>t transition mutations still occur in the non-coding strand before and after switching the reporter gene orientation, the result would provide significant support for the existence of TCR in euryarchaea. As mentioned previously, biochemical characterisation of the RNA polymerase of an archaeon, in terms of affinity for deaminated bases, would also provide support for a TCR pathway involving recognition of deaminated DNA. The most pressing issue, however, to result from this study concerns the need for a more robust reporter assay; one that provides a colourimetric indication of the status of the gene. If such a

reporter could be identified, the potential for revealing genetic studies of the archaea would be greatly improved.

Chapter 6

Deaminated base recognition and its effect on archaeal family D DNA polymerases

6.1: Background

The D family of DNA polymerases is perhaps the least well characterised of all polymerase classes and is found exclusively in the archaeal domain. Although originally identified in a euryarchaeon (Ishino *et al.*, 1998), sequence analyses have since revealed the presence of Pol D-encoding genes in the genomic sequences of korarchaea, nanoarchaea and thaumarchaea. However the enzyme is conspicuously absent from the crenarchaea. Pol D is invariably heterodimeric comprising a small subunit of approximately 70 kDa, on which the 3'-5' exonuclease domain resides, and a large subunit of approximately 140 kDa, which is responsible for polymerase activity (Ishino, 1998; Tang *et al.*, 2004). The enzyme has a primer-template preference that is characteristic of a replicative polymerase (Ishino *et al.*, 1998) and, based on its *in vitro* strand displacement activity, is believed to be responsible for lagging strand DNA synthesis (Henneke *et al.*, 2005). It has been proposed that Pol D may act soon after initiation by the primase and that, at a later stage, a switch occurs causing Pol B to take over replication of the leading strand while Pol D continues with lagging strand synthesis (Henneke *et al.*, 2005; Castrec *et al.*, 2009). As stated in chapter 5, the Pol D-encoding gene has been shown to be essential in a halophilic euryarchaeon, however, conclusive genetic evidence demonstrating the role of Pol D in chromosomal DNA replication is not currently available.

At present very little information exists regarding the ability of Pol D to recognise and respond to deaminated bases, as has been observed and well characterised with archaeal Pol B (Greagg *et al.*, 1999; Fogg *et al.*, 2002; Firbank *et al.*, 2008; Russell *et al.*, 2009). On the basis of a PCR assay, however, it has previously been reported that Pol D is incapable of incorporating dUTP into growing DNA strands and cannot replicate uracil-containing templates (Sawai *et al.*, 2007). DNA polymerases that specifically recognise and respond to deaminated bases are believed to have evolved and been maintained exclusively in the archaea due to their propensity for physically- and geochemically-extreme environments (Wardle *et al.*, 2008; Tahirov *et al.*, 2009). It is thought that the increased frequency of DNA deamination that occurs at high temperatures necessitates additional protective mechanisms to prevent the pro-mutagenic replication of uracil and hypoxanthine (Schouten and Weiss, 1999; Chen

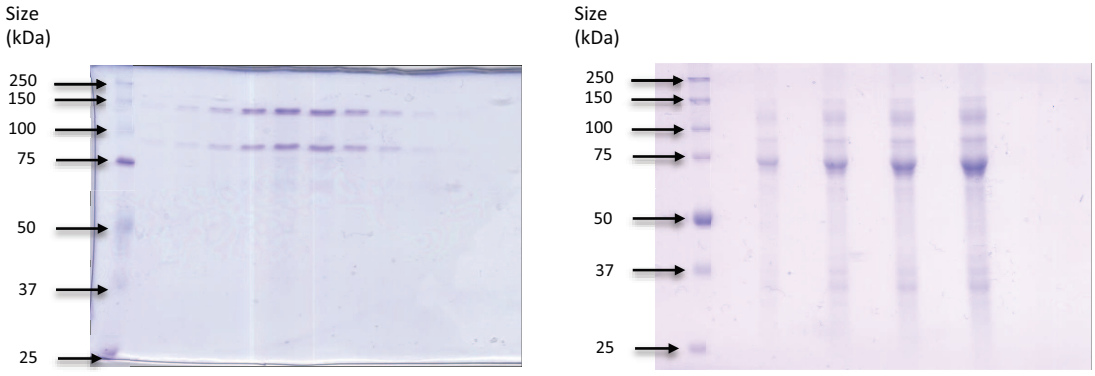
and Shaw, 1993; Schroeder and Wolfenden, 2007; Wardle *et al.*, 2008). Since Pol D is believed to function as a replicative polymerase and is thought to be responsible for copying ~50 % of the euryarchaeal genome (Ishino *et al.*, 1998; Henneke *et al.*, 2005), it seems likely that the enzyme would also possess specialised features to prevent replicative bypass of deaminated bases. However, at present no detailed investigation of deaminated base recognition by Pol D has been carried out and no specialised machinery, analogous to that of archaeal Pol B, has yet been identified.

In order to gain a greater understanding as to whether or not Pol D is capable of recognising deaminated bases and what physical features might be involved in this process, a comprehensive biochemical investigation has been carried out using the family D polymerases of *P. furiosus* and *P. abyssi*. The results of these analyses are presented throughout chapter 6.

6.2: Production of wild type and exo- Pfu-Pol D

Initially a number of primer extension, exonuclease, glycosylase and binding assays were performed using the family D polymerase of *P. abyssi*, which was generously provided by Dr Ghislaine Henneke of the IFREMER research institute in Brest, France. Concerns were raised, however, regarding the enzyme's long-term stability. SDS-PAGE analysis of the enzyme revealed significant degradation after transit to Newcastle (see figure 6.1b). As such, efforts were made to express and purify Pol D in Newcastle. Pfu-Pol D expression plasmids were generously donated by Professor Yoshi Ishino of Kyushu University, Japan. The small and large subunits of Pfu-Pol D were housed on the pET28a-PfuDP1 (His tagged) and pET22b-PfuDP2 (non tagged) plasmids, respectively. SDS-PAGE analysis and activity assays revealed that the enzyme from *P. furiosus* maintained long-term stability and retained activity for much longer periods than Pab-Pol D (figure 6.1). Amino acid sequence analysis of the two enzymes reveals that they share 86 % and 77 % sequence identity for the large and small subunits, respectively. Even when amino acid differences are found, in roughly 50 % of cases the substitutions are conservative. The behaviour of the two enzymes

was sufficiently similar in all of the assays presented in chapter 6 that the two could be used interchangeably. As such, some of the initial results presented here were produced using Pab-Pol D, whereas all of the latter work performed in the UK was done so using the enzyme from *P. furiosus*.



a) **b)**

Figure 6.1: 12 % SDS gels of the heterodimeric family D DNA polymerases of *P. furiosus* and *P. abyssi*. Panel **a)** shows peak fractions containing Pfu-Pol D after elution from a Superdex 200 10/300 GL column. Panel **b)** shows increasing amounts of purified Pab-Pol D. The amount of protein loaded in each lane ranges from 4-32 µg.

Based on the work of Gueguen *et al.*, 2001, a H445A mutation (corresponding to Pab residue H454) was introduced to the small, DP1, subunit of Pfu-Pol D to eliminate 3'-5' exonucleolytic activity. Exonuclease assays were performed to confirm the complete absence of hydrolytic degradation of DNA substrates (data not shown). The exo- enzyme was then used to perform binding titrations with a variety of DNA substrates, as described in chapter 6.3.

6.3: K_D determination of Pfu-Pol D by fluorescence anisotropy

Before performing binding titrations with Pfu-Pol D (exo-), the DNA duplexes that were to be tested (table 6.1) were subjected to gel shift assays confirming that the substrates remained fully annealed in the binding buffer that was later employed (figure 6.2). Although not shown in figure 6.2, the X-1 and double-stranded U:X substrates, which were synthesised at a later date, were also analysed by native PAGE analysis to confirm full annealing under experimental conditions.

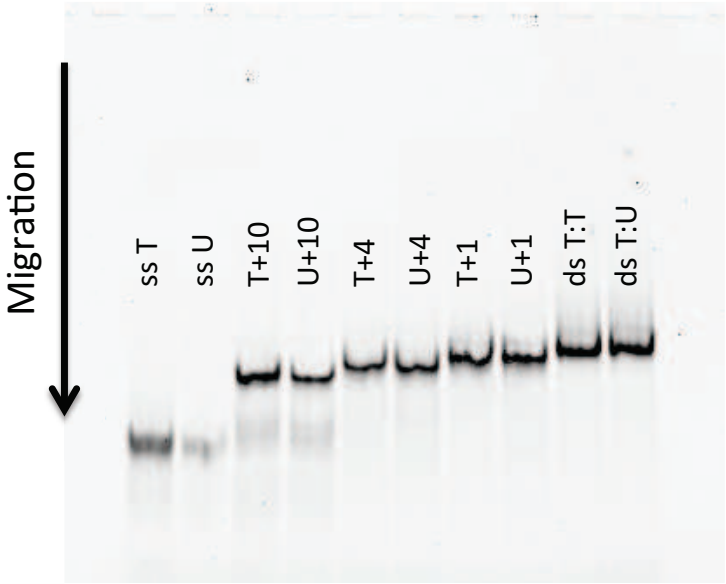


Figure 6.2: A mobility shift assay to assess hybridisation of Hex-labelled oligodeoxynucleotides by native PAGE. The full sequences of the DNA substrates are provided in table 6.3.

Fluorescence anisotropy was measured over a range of enzyme concentrations and the binding isotherms fitted by nonlinear least squares regression, as described by Reid *et al.*, 2001. Examples of the resulting binding curves are shown in figure 6.3. The plots were then used to calculate the binding affinities (K_Ds) for Pfu-Pol D with each of the tested substrates. The calculated K_Ds are summarised in table 6.1. Each K_D represents the mean of at least four titrations. The standard deviation is also provided for each value.

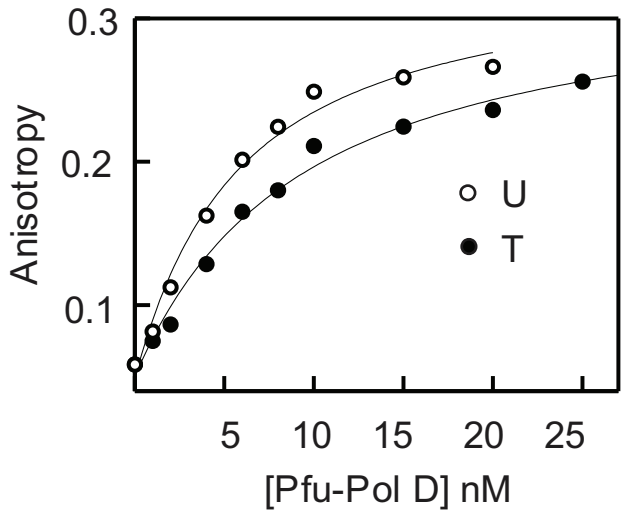


Figure 6.3: A representative example of the binding of Pfu-Pol D (exo-) to primer-templates containing dT or dU at the +4 template position (sequences given in table 6.1). The polymerase was added to the hexachlorofluorescein-labelled primer-templates and the increase in fluorescence anisotropy noted. The data was fitted to a 1:1 binding stoichiometry to give the titration curves shown. Binding constants are summarised in table 6.1.

DNA	**X** location	K_D (nM) X = T	K_D (nM) X = U
3′TTATCCAGGATATCCGCTTACCAGGTCGACC**X**TGGTCTTT-H*	single-stranded	6.0 ± 1.6	3.4 ± 0.3
5′AATAGGTCCTATAGGCGAATGGTCCAGCTGGAA 3′TTATCCAGGATATCCGCTTACCAGGTCGACC**X**TGGTCTTT-H*	-1	9.3 ± 0.6	5.2 ± 1.3
5′AATAGGTCCTATAGGCGAATGGTCCAGCTGG 3′TTATCCAGGATATCCGCTTACCAGGTCGACC**X**TGGTCTTT-H*	+1	9.1 ± 1.5	5.8 ± 0.4
5′AATAGGTCCTATAGGCGAATGGTCCAGC 3′TTATCCAGGATATCCGCTTACCAGGTCGACC**X**TGGTCTTT-H*	+4	7.3 ± 1.2	4.8 ± 0.3
5′AATAGGTCCTATAGGCGAATGG 3′TTATCCAGGATATCCGCTTACCAGGTCGACC**X**TGGTCTTT-H*	+10	8.4 ± 0.3	4.2 ± 0.5
5′AATAGGTCCTA**T**AGGCGAATGGTCCAGCTGGAACCAGAAA 3′TTATCCAGGATATCCGCTTACCAGGTCGACC**X**TGGTCTTT-H*	double-stranded	27.9 ± 6.9	29.0 ± 7.7
5′AATAGGTCCTA**U**AGGCGAATGGTCCAGCTGGAACCAGAAA 3′TTATCCAGGATATCCGCTTACCAGGTCGACC**X**TGGTCTTT-H*	double-stranded	29.4 ± 11.1	36.6 ± 2.4

Table 6.1: The K_D values for the binding of Pfu-Pol D (mean ± standard deviation from at least four determinations) to oligodeoxynucleotides containing uracil (or thymine in controls). The uracil residue is located at a variety of positions in both single and double-stranded DNA. The fluorophore, hexachlorofluorescein, is denoted as H* and was necessary to determine the K_D value by fluorescence anisotropy titration.

Pfu-Pol D was found to bind most tightly to single-stranded substrates, followed by primer-templates, then fully duplexed DNA. A preference of ~2 fold was observed for uracil-containing single-stranded and primer-template DNAs, compared to the equivalent controls. No significant preference was seen for any of the fully double-stranded substrates. Interestingly, the preference for uracil-containing primer-templates was retained when the uracil residue was located at the -1 position in the primer-template, relative to the primer-template junction, i.e. situated just within the double-stranded region.

6.4: Extension of uracil-containing primer-templates by Pfu-Pol D

Extension assays were performed using primer-templates containing uracil at a variety of positions, relative to the primer-template junction. Control primer-templates containing only the 4 canonical bases were also tested for comparison. The names and sequences of all of the substrates used in primer extension assays are shown in table 6.2.

Designation	Sequence
Control (gc/cg)	5′-Cy5-GGGGATCCTCTAGAGTCGACCTGC-3′ 3′-CCCCTAGGAGATCTCAGCTGGACGACCGTTCGTTCGAACAGAGG-5′
U+9 (gc/cg)	5′-Cy5-GGGGATCCTCTAGAGTCGACCTGC-3′ 3′-CCCCTAGGAGATCTCAGCTGGACGACCGTTCGUTCGAACAGAGG-5′
T+20 (gc/cg)	5′-Cy5-GGGGATCCTCTAGAGTCGACCTGC-3′ 3′-CCCCTAGGAGATCTCAGCTGGACGACCGTTCGTTCGAACAGAGTACCTGGCTAT-5′
U+20 (gc/cg)	5′-Cy5-GGGGATCCTCTAGAGTCGACCTGC-3′ 3′-CCCCTAGGAGATCTCAGCTGGACGACCGTTCGTTCGAACAGAGUACCTGGCTAT-5
Control (aa/tt)	5′-Cy5-GGGGATCCTCTAGAGTCGACCTGCAGGGCAA-3′ 3′-CCCCTAGGAGATCTCAGCTGGACGTCCCGTTCGTTCGAACAGAGG-5′
U-1 (aa/tt)	5′-Cy5-GGGGATCCTCTAGAGTCGACCTGCAGGGCAA-3′ 3′-CCCCTAGGAGATCTCAGCTGGACGTCCCGUTCGTTCGAACAGAGG-5′
U+4 (aa/tt)	5′-Cy5-GGGGATCCTCTAGAGTCGACCTGCAGGGCAA-3′ 3′-CCCCTAGGAGATCTCAGCTGGACGTCCCGTTCGTUCGAACAGAGG-5′

Table 6.2: Names and sequences of the substrates used in primer extension, dGTP incorporation and 3'-5' exonuclease assays in this chapter. Uracil (U) residues are highlighted in red.

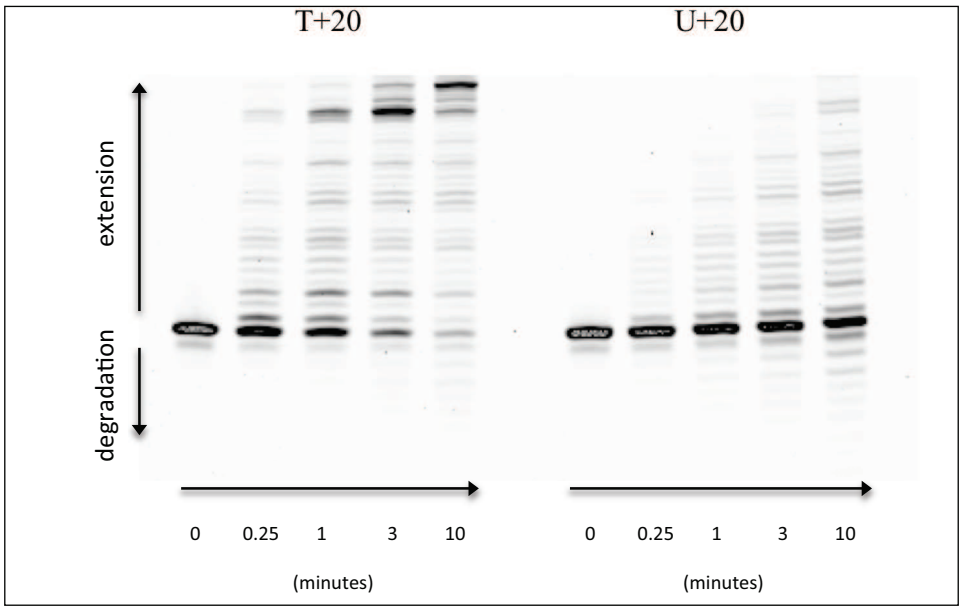


Figure 6.4: Primer extension assays performed using WT Pfu-Pol D with two primer-templates, the sequences of which are provided in table 6.2. Time points were taken at 0, 0.25, 1, 3 and 10 minutes.

The primer-extension assays shown in figure 6.4 reveal a significant level of inhibition of polymerisation when uracil is located at the +20 position, relative to the primer-template junction. Compared to the control reaction, considerably less fully extended material is seen by the final time point. A large amount of unextended primer is also visible at the end of the time course, as well as a significantly greater accumulation of degradation products. Very similar profiles of extension and degradation were observed with other sets of primer-templates where uracil was located at the -1, +4 or +9 positions (figure 6.5). Primer-templates lacking a g/c clamp were also tested to determine the effect of the terminal bases at the 3' end of the primer, however, once again very similar levels of inhibition of polymerisation were observed (figure 6.5). Compared to primer extension assays performed using archaeal Pol B, significant differences are immediately apparent. Unlike archaeal Pol B, there is no obvious position at which the family D polymerase stalls, relative to the deaminated base. Rather, a more general inhibition in polymerisation is seen, with the amount of extended product reduced at every position. Extension of uracil-containing primer-templates by Pol D also proceeds until fully extended primer is visible within a comparable time frame to the control. This is in stark contrast to extension assays performed using archaeal Pol B where virtually no fully extended product is seen following stalling, even after 24 hours (figure 3.4). In light of the findings presented in chapter 6.4, a variety of assays was employed to try and elucidate the mechanisms responsible for what appears to be a novel form of inhibition of polymerisation and to provide a more quantitative measure of the effects of uracil on Pol D.

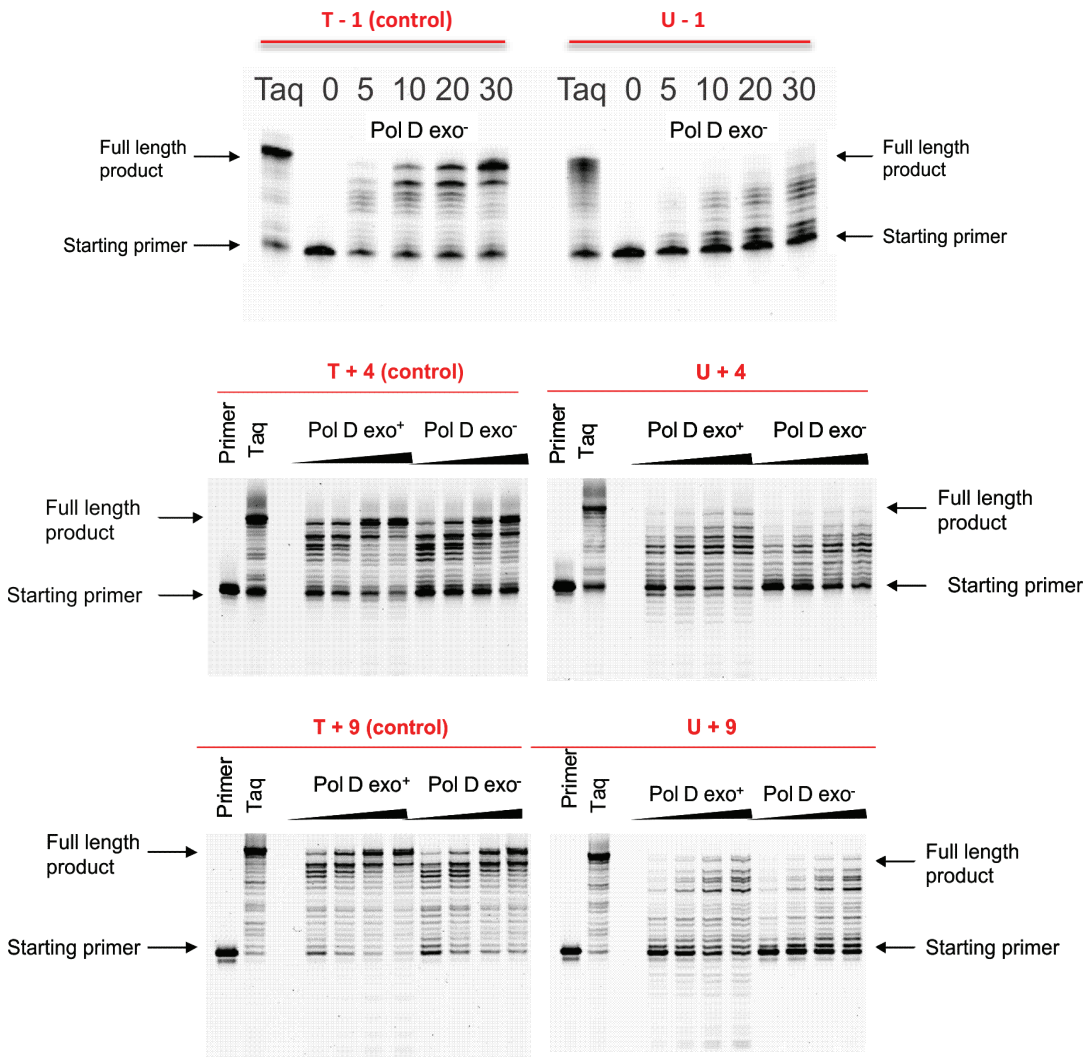


Figure 6.5: Primer extension assays performed using Pab-Pol D (wild type and exo-) with the X-1, X+4 and X+9 primer-templates described in table 6.2. In all of the assays time points were taken at 0, 5, 10, 20 and 30 minutes

6.5: dUTP incorporation assays

Following the observation that Pol D is incapable of producing PCR products when dUTP is substituted for dTTP (Sawai *et al.*, 2007), extension assays were performed to establish exactly how dUTP inhibits polymerisation. Two extension assays were performed using the Control (aa/tt) primer-template described in table 6.2. One

reaction mix contained the four canonical dNTPs, while the other contained dUTP in place of dTTP. The results of the two assays are shown in figure 6.6.

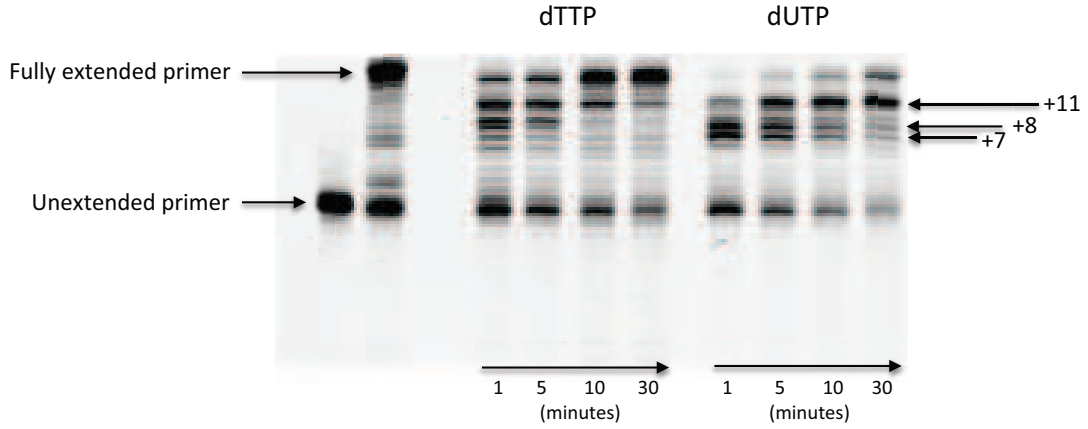


Figure 6.6: Comparison of dTTP and dUTP incorporation by Pab-Pol D into the control (aa/tt) primer-template described in table 6.2. The unextended primer is shown in the lane farthest to the left. A fully extended primer, produced by Taq polymerase, is shown in the adjacent lane.

As with the extension assays described in chapter 6.4, a mild inhibition in polymerisation was observed with the dUTP reaction. Less fully extended product is visible in the dUTP reaction by the final time point. A small but significant pause is seen around positions +7, +8 and +11, approximately corresponding to regions where dUTP incorporation would be expected to occur. Since the +7 and +8 template strand bases are both adenine, the results suggest that dUTP incorporation occurs with the same efficiency as dTTP but, once incorporated, uracil appears to inhibit extension of the primer strand. Surprisingly, unlike the assays described in chapter 6.4, there appears to be a reduction, rather than an increase, in the amount of 3'-5' exonucleolytic degradation observed with the dUTP assay, compared to the control.

6.6: Uracil-DNA glycosylase assays

In order to determine whether or not Pol D might exhibit a previously unidentified uracil-DNA glycosylase activity, a number of glycosylase assays was performed using exonuclease-proficient and -deficient strains of Pab-Pol D. Enzymes were incubated with a variety of DNA substrates (table 6.3) in the absence of dNTPs. The results of these assays are shown in figure 6.7.

Designation	Sequence
ss T	5'-Hex-AATAGGTCGATATCGCGAATGG-3'
ss U	5'-Hex-AATAGGTCGATAUCGCGAATGG-3'
ds TT	5'-Hex-AATAGGTCCTATAGGCGAATGG-3' 3'-TTATCCAGGATATCCGCTTACC-5'
ds TU	5'-Hex-AATAGGTCGATAUCGCGAATGG-3' 3'-TTATCCAGCTATAGCGCTTACC-5'
ds UU	5'-Hex-AATAGGTCGATAUCGCGAATGG-3' 3'-TTATCCAGCUATAGCGCTTACC-5'

Table 6.3: Names and sequences of substrates used in uracil-DNA glycosylase assays of Pab-Pol D. Uracil (U) residues are highlighted in red.

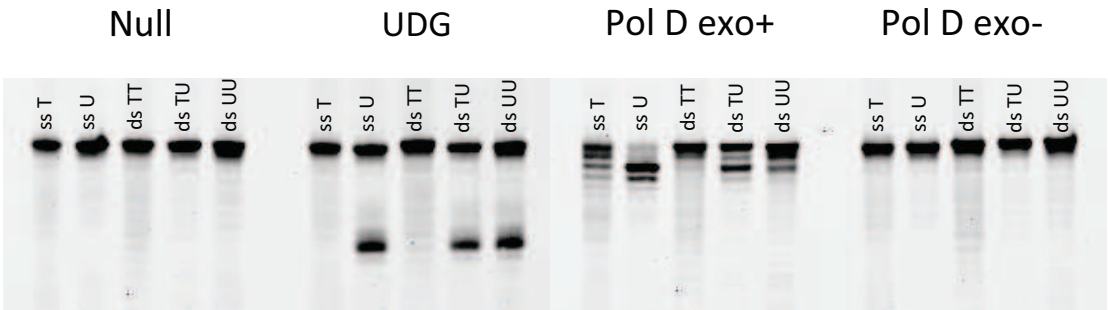


Figure 6.7: Denaturing PAGE analysis of the products of uracil-DNA glycosylase assays performed using the DNA substrates shown in table 6.2 with Pab-Pol D (exo+ and exo-). Null and UDG controls were also performed.

Figure 6.7 confirms that the null reactions had no effect on the substrates and that all five oligodeoxynucleotides migrate at a similar rate when analysed by denaturing PAGE. The UDG controls show that the uracil-containing Hex-labelled substrates are all acted upon by UDG to produce an abasic (AP) site. The AP site was then assumed

to be cleaved during heating of the samples to form the truncated product observed with the ss U, ds TU and ds UU substrates. None of the Pab-Pol D assays formed products that would indicate a UDG-type activity. Similar assays were also performed using primer-template substrates, however, once again no glycosylase activity was observed with Pol D.

6.7: Single turnover dGTP incorporation assays

In order to produce a more quantitative measure of Pfu-Pol D's ability to replicate uracil-containing primer-templates, single nucleotide incorporation assays were performed using primer-templates containing either thymine or uracil at the +4 position, relative to the primer-template junction. This approach is widely used with DNA polymerases (Joyce, 2010) and has previously been made use of to investigate the interaction of archaeal family B polymerases with deaminated bases (Emptage *et al.*, 2008). A H445A (exo-) version of Pfu-Pol D was employed to enable investigation of just the polymerising activity of the enzyme. Preliminary results showed that at 20 nM concentrations of primer-template, 140 nM levels of Pfu-Pol D were enough to ensure complete DNA binding, hence reactions were carried out under genuine single turnover conditions (data not shown). This is expected given the affinities for DNA reported in table 6.1 but differs from archaeal Pol B, which exhibits weak binding with control primer-templates and requires PCNA to ensure saturation (Emptage *et al.*, 2008). The addition of 5 to 400 μM levels of dGTP to Pfu-Pol D and primer-template resulted in the extension of the primer by a single base, as assessed by gel electrophoresis (figure 6.8), and permitted determination of the rate of the observed rate of incorporation, k_{obs}, by fitting the amount of extended primer against time to a single exponential (not shown). A secondary plot of k_{obs} versus dGTP concentration, fitted using the Michaelis-Menten equation, gave the results shown in figure 6.9 and the kinetic constants summarised in table 6.3. From the kinetic constants, it is clear that the presence of uracil inhibits polymerisation almost entirely by reducing k_{pol}. k_{pol} is reduced by a factor of ~5 for the uracil-containing primer-template with little change in the K_D for dGTP.

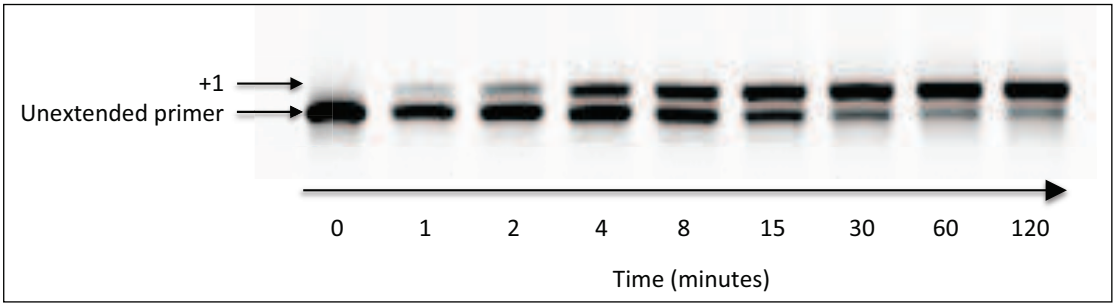


Figure 6.8: An example of a single turnover dGTP incorporation assay using Pfu-Pol D exo-. In this case the primer-template used was U+4 (aa/tt) (described in table 6.2) with 10 μM dGTP.

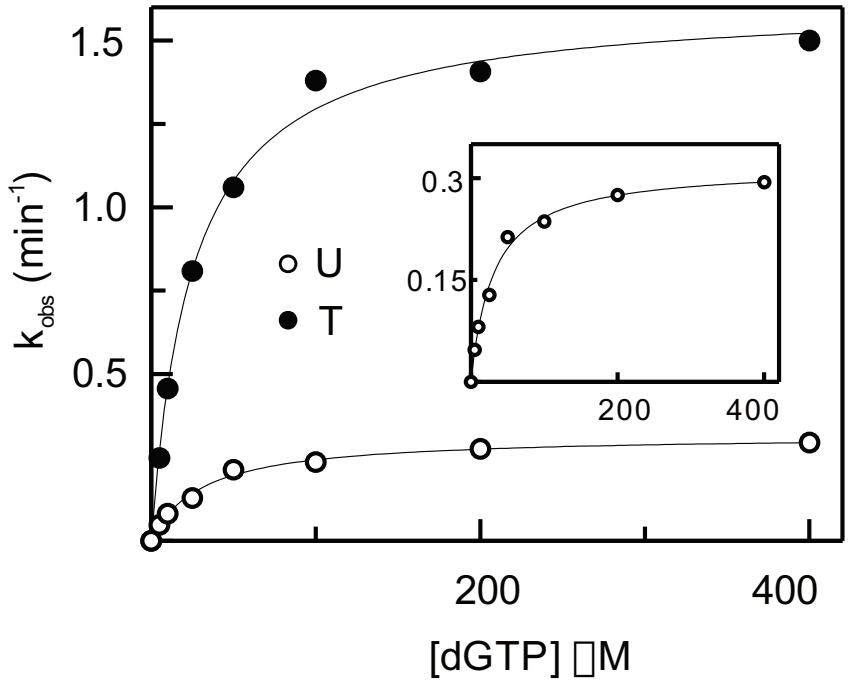


Figure 6.9: Incorporation of a single dGTP into primer-templates containing dT or dU at the +4 template position (sequences given in table 6.2). The primer-templates (20 nM) were mixed with Pfu-Pol D (140 nM) and the reaction initiated by adding appropriate amounts of dGTP (between 5 and 400 μM) and the amount of primer extended by a single base at various times was determined by gel electrophoresis (not shown). This data was used to determine a k_{obs} value for each concentration of dGTP (not shown). Secondary fits to the Michaelis-Menten equation are given for both the dT and dU containing primer-templates (the insert is an expansion of the dU data). The kinetic parameters determined from this graph are summarised in table 6.4.

base at +4 position of primer-template[a]	k_{pol}[b] (min^{-1})	K_D[b] (μM)	k_{pol}/K_D[b] (s^{-1} M^{-1})
dT	1.6 ± 0.2	25 ± 6	3.8 ± 1.4 x 10^6
dU	0.3 ± 0.06	31 ± 9	0.6 ± 0.3 x 10^6

Table 6.4: Kinetic parameters for incorporation of a single dGTP into a primer-template containing either dU or dT at +4 under single turnover conditions.

[a]The primer-template used in this experiment was:

```
5'Cy5-GGGGATCCTCTAGAGTCGACCTGCAGGGCAA-3'
   3'-CCCCTAGGAGATCTCAGCTGGACGTCCCGTTCGTXCGAACAGAGG-5'
```

Where X = dU or dT

[b]Kinetic parameters are the means (± standard deviation) from five experiments.

6.8: Pfu-Pol D 3'-5' exonuclease assays

The extension assays presented in chapter 6.4 show that a small increase in 3'-5' exonucleolytic activity was observed with uracil-containing primer-templates. In order to gain a more quantitative measure of the effect of template strand uracil on the 3'-5' exonucleolytic activity of Pfu-Pol D, single turnover exonuclease assays were performed using two primer-templates. The primer-templates contained either uracil or thymine at the +9 position, relative to the primer-template junction. Attempts were made to measure the 3'-5' exonucleolytic activity of Pfu-Pol D using the control and U+4 (aa/tt) primer-templates employed in chapter 6.7, however degradation of the primer-templates proceeded too quickly to be measured by hand and instead required stopped-flow analysis. As such, the control and U+9 (gc/cg) primer-templates were employed instead. As with the dNTP incorporation experiments, proofreading exonuclease rates were measured with 20 nM concentrations of primer-template and 140 nM Pfu-Pol D. The results of the assays can be seen in figure 6.10 and reveal that the presence of uracil at the +9 position results in a ~3-fold increase in the rate of 3'-5'exonucleolysis exhibited by Pfu-Pol D, as compared to the rate observed for the control reaction.

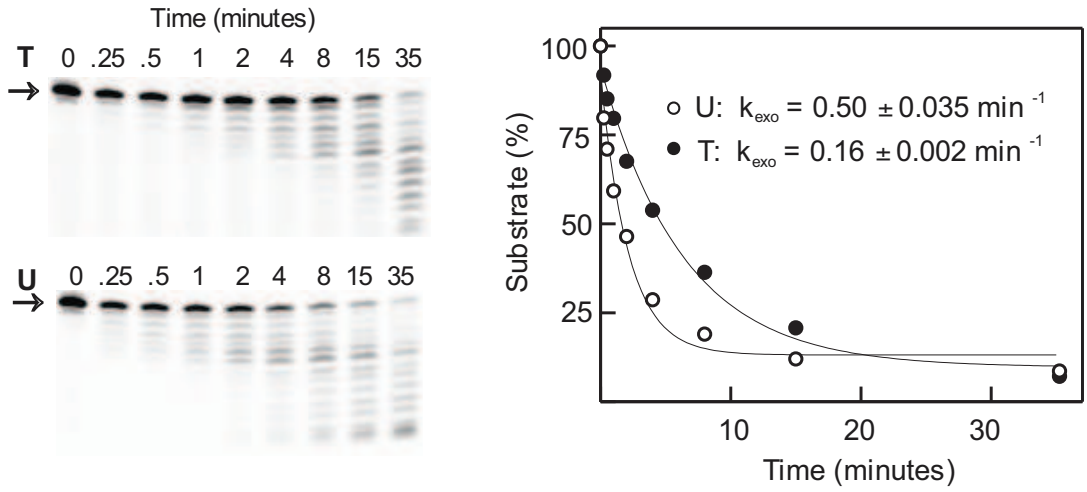


Figure 6.10: Proofreading exonucleolysis of primer-templates containing dT or dU (sequences given below) by Pfu-Pol D. The polymerase (140 nM) was added to the primer-templates (20 nM) and the degradation of the Cy5-labeled primer noted for both dT (gel labeled T) and dU (gel labeled U) containing templates. The gels observed are shown, with the starting primers marked with an arrow. Also given is a fit to a single exponential of the amount of substrate remaining with time, which was used to determine the rate constants for exonucleolysis. The k_{exo} values found are shown on the graph and are the averages (± standard deviation) from three experiments. The primer-templates used in these experiments were:

```
5'-Cy5-GGGGATCCTCTAGAGTCGACCTGC-3'
    3'-CCCCTAGGAGATCTCAGCTGGACGACCGTTCGXTCGAACAGAGG
```

where X = dT or dU.

6.9: "Masking" of template strand uracil prevents specific recognition by Pol D

The data presented in chapter 6.9 were all produced by Dr Ghislaine Henneke and colleagues using Pab-Pol D.

In order to test whether or not Pol D is still capable of recognising uracil when the deaminated base is buried deeply in double-stranded regions of DNA primer-templates, primer-extension assays were performed using an alternative configuration of substrate. A Cy5-labelled 17mer primer was annealed to an 87mer template containing either cytosine or uracil at the +16 position, relative to the primer-template junction (figure 6.11A). The results of primer extensions performed using these substrates indicate that polymerisation by Pab-Pol D is inhibited by the presence of uracil, as previously observed. A similar 17/87mer configuration of primer-template was then employed but with uracil or thymine located at the +59 position, relative to the primer-template junction (figure 6.11B). As before, when uracil was located in the single-stranded region of the template an inhibition of polymerisation was observed. However, when a "masking" oligodeoxynucleotide was employed, annealing to the single-stranded template region of the substrate to mask the uracil residue, thereby creating a gapped substrate, the inhibiting influence of the deaminated base was completely lost.

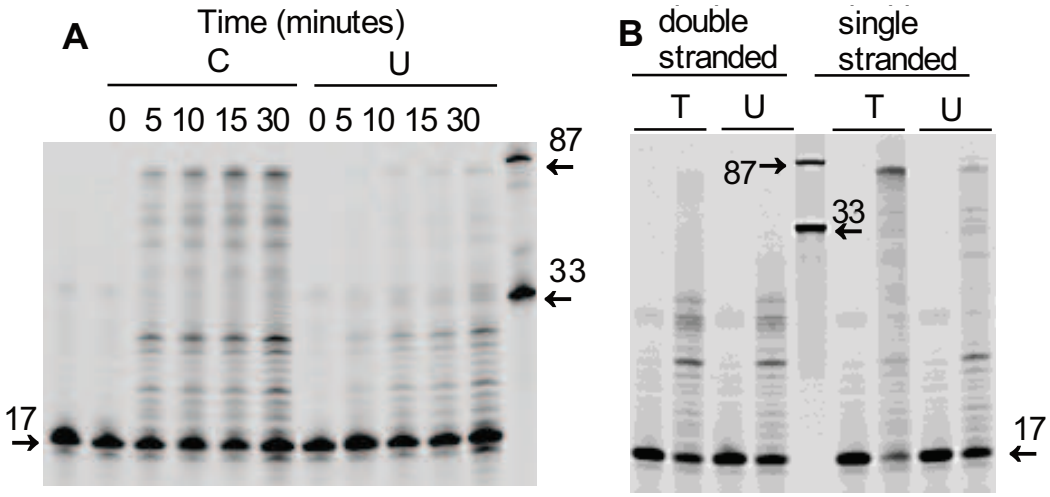


Figure 6.11: Inhibition of Pab-Pol D by template strand uracil. **A:** Extension of 17/87 primer-templates (sequences given below) by Pab-Pol D. The image shows the amount of extended product formed at various times with dC (labeled C on the gel) or dU (labeled U on the gel) at position +16 in the template. Reference oligodeoxynucleotides of 17, 33 and 87 bases are indicated by the arrows. **B:** Influence of burying uracil in a double-stranded region of the substrate. 17/87 primer-templates containing either dT (T) or dU (U) at +59 with or without a "masking" oligodeoxynucleotide (labelled single or double-stranded) were extended by Pab-Pol D. Reference oligodeoxynucleotides of 17, 33 and 87 bases are indicated by the arrows.

The 17/87 primer-templates used for panels A and B were:

```
5′-Cy5-TGCCAAGCTTGCATGCC-3′          5′-GAGGATCCCCGGGTACCGAGCTCGAATTCGTAATCATGGTCATAGCTGTTTCCTG-3′
    3′-ACGGTTCGAACGTACGGACGTCCAGCTGAGATXTCCTAGGGGCCCATGGCTCGAGCTTAAGCATTAGTACCAGTAYCGACAAAGGAC-5′
```

Panel A: A duplex consisting of the 17-mer primer annealed to the 87-mer template was used. **X** = dC or dU; **Y** = dT. The third "masking" strand was not used in these experiments.

Panel B: An additional "masking" oligodeoxynucleotide was added to locate uracil at +59 in a double-stranded region. In controls this strand was missing, placing dU+59 in a single-stranded region. **X** = dC; **Y** = dT or dU.

6.10: Probing the role of Pfu-DP2 C953 in uracil recognition

Due to the lack of X-ray crystal data for the whole Pol D enzyme, it is currently difficult to speculate as to which physical features might be facilitating the specific recognition of uracil reported in chapters 6.3-6.9 of this thesis. Analysis of the amino acid sequence of the large subunit of Pol D, however, reveals an unusually large number of highly conserved cysteine residues, almost all of which are situated in pairs (figure 6.12). Based on current knowledge of protein-metal affinity the observed profile of amino acids in Pol D could be indicative of metal-binding sites, such as iron-sulphur clusters (Otvos and Armitage, 1980; Passerini *et al.*, 2006). At present, however, there is no direct evidence to support this hypothesis.

Type 1:

Type 2:

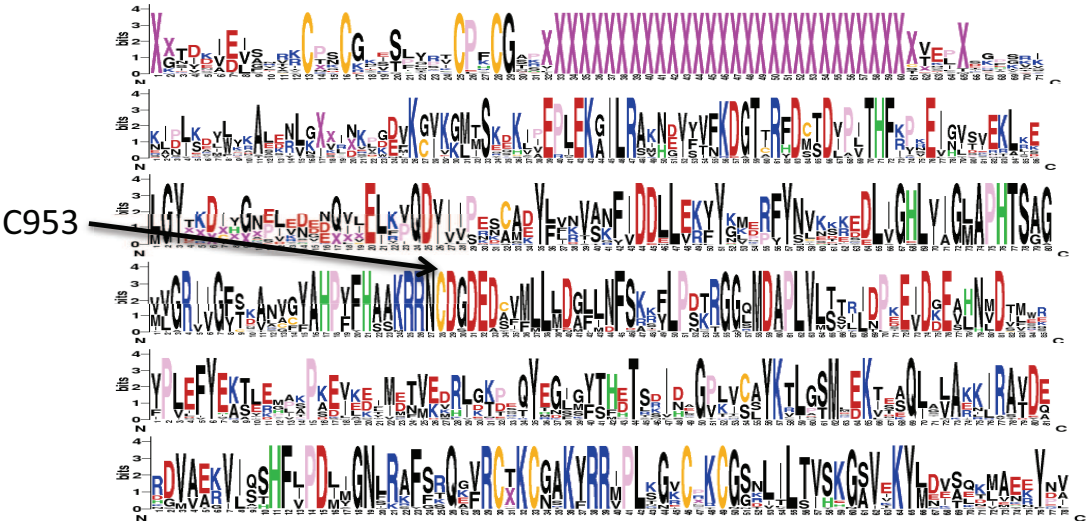
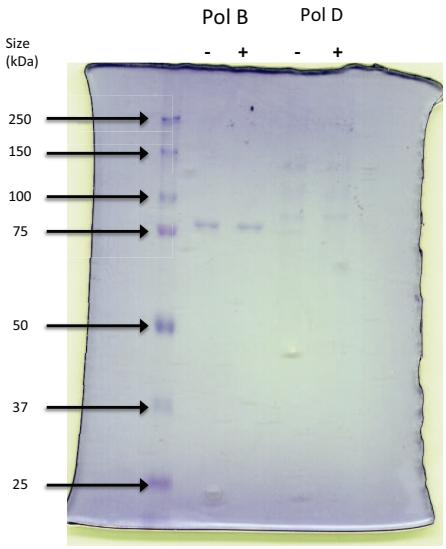


Figure 6.12: Amino acid sequence conservation of the large subunit C-terminus of family D DNA polymerases. The degree of conservation of each residue is indicated by the size of each letter. Generally speaking, sequences can be categorised as belonging to either type 1 or type 2. The two are almost identical, except for a region containing 30 amino acids, which is absent from the type 2 sequence and is indicated by the string of magenta Xs on line 1. Family D polymerases from the *Pyrococcus* and *Thermococcus* genera are invariably type 1, while family D polymerases from the other genera can be classed as type 2 (Image produced using weblogo [http://weblogo.berkeley.edu/])

As well as the unusual number of paired cysteines present in Pol D, it was noted that one residue, C953, is located entirely on its own and was present in all of the analysed sequences. Owing to its unusual location and high degree of conservation, it was proposed that the residue might be playing a fundamental role in stabilising the enzyme, or facilitating a key mode of action of the enzyme. One hypothesis was that the thiol group of C953 might form of a stable covalent adduct with uracil, as has been observed with RNA-binding proteins (Lin *et al.*, 2003). In order to test this theory, the reactive nucleotide analogue 2-pyrimidinone was employed. 2-pyrimidinone forms reasonably stable covalent adducts with cysteine residues in DNA methyltransferases (Subach *et al.*, 2004). So in order to test whether or not C953 or any of the other cysteine residues are involved in DNA binding, Pfu-Pol D exo- was incubated with a single-stranded DNA substrate containing 2-pyrimdinone and analysed by SDS-PAGE. Unfortunately, SDS-PAGE analysis of the protein-DNA complexes showed no shift in the size of either subunit (figure 6.13). A C953A mutation was also introduced to the enzyme to investigate the effect of substituting the residue on the polymerising activity of Pfu-Pol D (figure 6.14). Control and uracil-containing primer-templates were employed in primer extension assays, revealing a decrease in polymerising activity by the mutant enzyme. However, a similar reduction in polymerisation was seen with both substrates and the control reaction still proceeded significantly faster than the uracil.



single-stranded 2-pyrimidinone: 5'-TAAGCGAP^{yr}ATCCTGAT-3'

Figure 6.13: SDS PAGE analysis of Pfu-Pol B and Pfu-Pol D after incubation with a single-stranded oligodeoxynucleotide, the sequence of which is shown within the figure. P^{yr} denotes pyrimidinone. The enzyme on its own is indicated by the **"-"** symbol. The enzyme and DNA substrate is indicated by the **"+"** symbol.

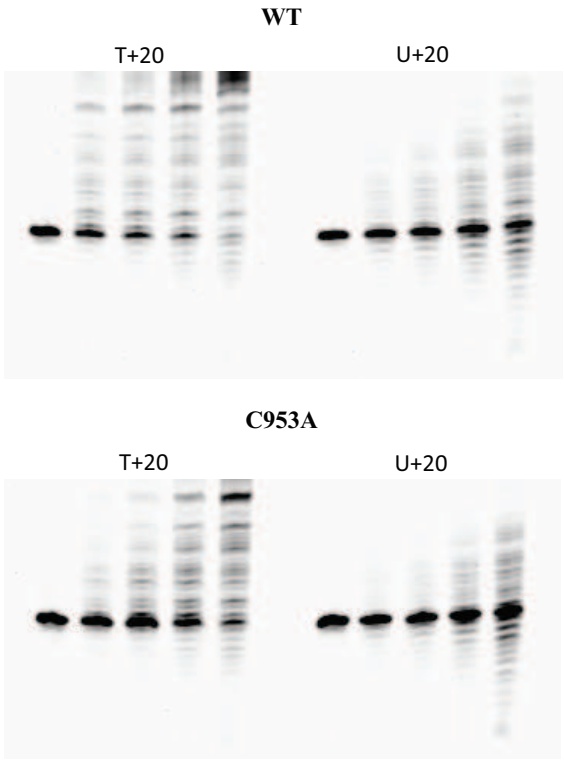


Figure 6.14: Extension of T+20 and U+20 primer-templates by Pfu-Pol D WT and C953A. The full sequences of the primer-templates are given in table 6.2.

6.11: Discussion

The data presented in this chapter show that the occurrence of uracil in DNA diminishes polymerisation by euryarchaeal family D polymerases. Uracil must be positioned in the single-stranded template region of primer-templates or near the primer-template junction of double-stranded regions for the inhibitory effect to be observed. When the residue is embedded well within duplex DNA little or no inhibition is observed. Therefore it seems likely that Pol D is only able to specifically interact with uracil in single-stranded DNA, but may capture the base in duplexes following transient melting of terminal regions. Preferential binding of uracil by Pol D is also only observed when the base is situated in a single-stranded context, agreeing with the same requirement for inhibition of polymerisation. When uracil is present in single-stranded templates small changes to a number of parameters are observed: binding affinity is increased by a factor of ~2, polymerisation decreases

about five-fold (as measured using k_{pol}) and the 3′-5′ proofreading exonuclease rates are about three times faster (as measured by k_{exo}). Together these alterations result in markedly less DNA synthesis by Pol D when uracil is present in template strands, as compared to controls. Since DNA polymerisation involves addition of multiple dNTPs, a relatively small change in the efficiency of incorporation for each single base (such as the 5-fold reduction in k_{pol} seen in this study) may cumulatively lead to profound overall inhibition.

Template strand uracil has previously been observed to strongly suppress the activity of archaeal family B polymerases, a feature assumed to reduce mutations that arise as a consequence of replicating deaminated cytosine (Greagg *et al.*, 1999; Fogg *et al.*, 2002; Firbank *et al.*, 2008; Killelea *et al.*, 2010). It is most likely that the uracil-dependent inhibition of Pol D serves an identical function in protecting the genome from base deamination. The ability of the two polymerases to detect and respond to uracil suggests they both play a significant role in chromosomal DNA replication, as has been previously suggested (Henneke *et al.*, 2005; Castrec *et al.*, 2009; Rouillon *et al.*, 2007). However, biochemical studies can only tentatively identify a polymerase as replicative and the exact roles of Pol B and Pol D in replication of the euryarchaeal genome await genetic investigation. Interestingly, the mechanisms by which the two polymerases interact with uracil appear to be quite distinct from one another. The mode of interaction of Pol B has been well characterised, involving "read-ahead recognition", where a running polymerase scans the template ahead of the replication fork for the presence of uracil and senses the base via an N-terminal pocket (Greagg *et al.*, 1999; Fogg *et al.*, 2002; Firbank *et al.*, 2008; Killelea *et al.*, 2010). The hallmarks of "read-ahead recognition" are strong and specific binding of uracil and stalling of replication at a highly defined position, four bases prior to uracil encounter. In the unlikely event that replicative bypass of uracil occurs, inhibition of Pol B ceases. Pol D shows none of these features and therefore must use a novel mechanism to sense uracil. Binding of the deaminated base is far from profound and a unique truncated extension product, which would suggest stalling at a defined position, is not seen. Instead, all of the bands that represent polymerisation products, as detected by denaturing PAGE analysis, are diminished; particularly those of longer products. Furthermore, Pol D is able to interact with template strand uracil at positions well beyond the primer-template junction. Interaction is even seen when

uracil is situated in double-stranded regions of DNA, just behind the fork. Should Pol D progress beyond uracil, inhibition transiently persists until further replication positions the base deep within the duplex. Once buried deeply with the double-stranded DNA, melting to produce single strands becomes less likely. The DNA repair processes that follow uracil sensing by both Pol B and Pol D await elucidation but must involve the accurate replacement of uracil with cytosine, the parent base from which it is derived by deamination. The near complete halting of replication when Pol B encounters uracil represents a very efficient means of buying time for downstream repair to occur. The slowing of polymerisation, characteristic of Pol D, seems at first sight to be a less effective strategy. However, the ability of the enzyme to recognise uracil well in advance of the replication fork presumably ensures success. The continuing inhibition following the copying of uracil is counterintuitive, as the damage (i.e. incorporation of adenine opposite uracil) has already been done. However, as previously mentioned, the inhibition in polymerisation fades as extension progresses.

The molecular mechanism that gives rise to the uracil-dependent inhibition of replication seen with Pol D is far from clear and a tentative proposal is given in figure 6.15. Two polymerase-DNA complexes are suggested to exist in equilibrium with each other. One species (labelled A in figure 6.15) forms the standard interactions expected of a polymerase with the primer-template junction of the DNA substrate. This species is assumed to be efficient at extension. In the second complex (labelled B) an additional interaction is seen with uracil, resulting in a conformational change that inhibits the polymerase active site, thereby slowing the rate of replication. Conformation B also appears to exhibit a higher rate of proofreading exonuclease activity. The recognition of uracil, regardless of its template strand location, can be accounted for by looping out the single-stranded DNA. At present it is unclear whether or not form B completely lacks the ability to extend DNA substrates. If so, residual activity must arise from the remaining conformation A in the equilibrium. Alternatively, species B may have reduced extension ability. In this case, shifting the conformational equilibrium 100 % towards this form will still not completely abolish polymerisation activity. It should be noted that a uracil-dependent complete cessation of polymerisation was never observed in contrast to an earlier brief publication, which reports full inhibition of Pol D by uracil (Sawai *et al.*, 2007). Pol D is a heterodimer

(Cann *et al.*, 1998; Cann and Ishino, 1999) but it is presently unknown if the uracil sensing apparatus is located in the large polymerase subunit or the smaller exonuclease subunit. The heterodimer has been reported to further assemble into a tetramer (Shen *et al.*, 2003). However, it is unresolved whether a single heterodimer interacts with both the primer-template junction and uracil, or if two separate heterodimers present in the tetramer are used for individual recognition events.

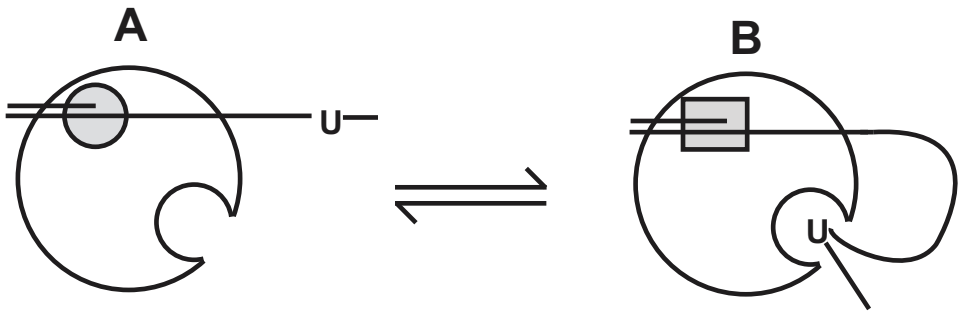


Figure 6.15: Model for the interaction of Pol D with uracil. Conformation A is found in the absence of uracil, with the primer-template binding active site (grey circle) having high polymerisation activity. Uracil, in single-stranded templates, can bind at a remote site to give conformation B. Here the active site (grey square) shows a lowered extension rate. This form also has increased 3′-5′ proofreading exonuclease activity. Alternatives are discussed in the conclusions section.

One concern about the model shown in figure 6.15 is that the proposed uracil recognition site may be anticipated to result in a binding selectivity higher than the ~2-fold increase observed in this study. The uracil-binding pocket of archaeal Pol B, for example, increases affinity for DNA substrates that contain uracil 200-300-fold, compared to the affinity shown for equivalent substrates that contain only the four canonical bases (Shuttleworth *et al.*, 2004; Gill *et al.*, 2007). The K_D values given in table 6.1 are, however, unlikely to represent the true or intrinsic binding constant for the interaction with uracil. Pol D is expected to interact with a particular length of single-stranded DNA in a number of binding modes. As such, the K_D values seen with controls represent the sum of every individual non-specific binding constant; those with uracil arise from non-specific binding plus any specific interaction with

uracil. Therefore, although the K_Ds shown in table 6.1 are apparent values, the small but significant preference for uracil in single-stranded regions is clear.

At present no high-resolution crystal structure is available for the entire Pol D, although information is available for both the N-terminal regions of the small and large sub-units (Yamasaki *et al.*, 2010; Matsui *et al.*, 2011). Unfortunately, inspection does not reveal any obvious uracil-binding pocket, analogous to that of archaeal Pol B (Firbank *et al.*, 2008; Killelea *et al.*, 2010). As stated in chapter 6.9, however, Pol D contains a large number of cysteine residues arranged in pairs, suggesting it may exist as an iron-sulphur protein, as observed for two archaeal DNA repair enzymes, uracil-DNA glycosylase (UDG) and the repair helicase, XPD (Hinks *et al.*, 2002; Rudolf *et al.*, 2006). The "single" cysteine, C953, does not appear to form a transient covalent adduct with uracil and is therefore unlikely to account for specific binding. It also appears to play little or no role in uracil recognition as mutagenesis of C953 to alanine results in the production of a polymerase that retains uracil-sensitivity. The samples of Pol D used in this investigation were recombinant; they were overexpressed in *E. coli* and purified under aerobic conditions, never exhibiting the characteristic olive green hue of iron-sulphur clusters. Thus it is possible that the proposed cluster(s), discussed in chapter 6.10, may have been lost and that correct folding of the uracil-sensing region may not have been achieved. Further studies are currently in progress, including the preparation of Pol D under native, anaerobic conditions to maintain any putative iron-sulphur clusters. It is hoped that this approach should facilitate a more thorough characterisation of the unusual and novel interaction that this protein exhibits with deaminated bases.

Chapter 7

Summary

7.1: Achievements

During the course of this PhD significant progress has been made in improving our current understanding of the deaminated base recognition pathways of archaeal DNA polymerases. The events that follow replicative stalling by archaeal Pol B upon encountering template strand uracil or hypoxanthine has been elucidated. The results presented in chapter 3 conclusively show that Pfu-Pol B exhibits a 50-200 fold increase in 3'-5' exonucleolytic activity when approaching template strand deaminated bases. This stimulation occurs when the enzyme is located between 0 and 3 bases upstream from uracil or hypoxanthine. When located at these positions the polymerase activity of the enzyme is simultaneously inhibited. Modulation of the enzyme's two primary activities serves to maintain the polymerase at a preferred position, 4 bases upstream from the deaminated base. It has been proposed that maintaining this position serves to delay DNA replication long enough to enable an as yet uncharacterised repair pathway to take place. However, at present, the mechanisms involved in the proposed repair pathway remain unknown.

The manner in which archaeal Pol B achieves this stimulation in 3'-5' exonucleolytic activity has also been identified. From the results of fluorescence studies using the nucleotide analogue, 2AP, unwinding and separation of the primer and template strands appears to be the underlying mechanism for stimulation of 3'-5' exonucleolytic activity by Pfu-Pol B. Two amino acids, M247 and Y261, contained within a β hairpin region of the enzyme have been identified as playing a key role in DNA strand partitioning by site directed mutagenesis. The exact nature of the relationship between DNA strand partitioning and proofreading exonucleolysis by archaeal Pol B, however, remains unclear.

By genetic manipulation of the archaeon, *Thermococcus kodakarensis*, the role of deaminated base recognition by Pol B in maintaining genomic stability has been studied *in vivo* for the first time. Deaminated base recognition appears to contribute to maintaining genomic stability by preventing c>t transition mutations, possibly in a strand-specific manner. The extreme bias in the rate of c>t transition mutations in favour of the non-coding strand raises the intriguing possibility of a transcription-

coupled repair pathway in an archaeon. The potential link to deaminated base recognition is also of interest. However, further investigation is required to improve the quality and significance of the data and establish if either of these possibilities is likely to be the case.

During the course of this thesis the ability of family D DNA polymerases to recognise and respond to deaminated bases has been investigated in detail for the first time. While much subtler than the effect seen with archaeal family B polymerases, deaminated bases do appear to induce several small but significant changes in the kinetic parameters exhibited by Pol D. Polymerising activity appears reduced when template strand uracil is encountered in single-stranded template DNA, regardless of its distance from the primer-template junction. Similarly, a small increase in 3'-5' exonucleolytic activity is exhibited by Pol D when uracil is encountered. The underlying mechanism for this response is not yet fully understood but appears to result, at least partly, from the higher affinity of Pol D for uracil-containing substrates. It is proposed that distinct binding sites are responsible for uracil recognition and polymerase/exonuclease activity and that binding at the uracil recognition site results in the adoption of an alternative conformation, which is inhibited for polymerising activity. This second conformation also appears to exhibit higher rates of 3'-5' exonucleolysis.

7.2: Future work

In terms of the work on archaeal Pol B, several key questions remain regarding the deaminated base recognition and repair pathway. Firstly, as discussed in chapter 4, the exact relationship between primer-template strand separation, as occurs when archaeal Pol B approaches uracil or hypoxanthine, and 3'-5' exonucleolytic activity remains to be conclusively shown. Stopped flow fluorescence, to establish the rates at which the conformational change occurs, would be useful in extrapolating the exact sequence of events and the relationship between primer degradation and strand partitioning. Perhaps most critically of all, the downstream steps that follow deaminated base stalling by Pol B remain unknown. It seems likely that stalling

serves to enable uracil to be substituted for cytosine and it has previously been suggested (Greagg *et al.*, 1999) that the repair pathway might involve a recombination event. However, it is perhaps more likely that the uracil residue is instead excised from the exposed DNA by the action of a uracil-DNA glycosylase and subsequently replaced by cytosine. In order to identify the possible partners in this process, a genetic approach is likely to be fruitful. By employing an archaeal strain containing a His-tagged family B DNA polymerase, it should be possible to "fish out" a stalled complex. The tagged protein could be immobilised on a HisTrap column and any interacting partners eluted, precipitated and identified by mass spectrometric analysis. The likelihood of encountering a stalled Pol B complex could be improved through the employment of knockout strains of UDG or dUTPase, thereby promoting the build-up of genomic uracil.

Regarding the genetic study of *T. kodakarensis*, a much more robust reporter assay, equivalent to the commonly used system based on β-galactosidase and X-gal, is required before further investigation can be undertaken and firm conclusions made. Provided such a reporter can be identified, the potential for elucidating the respective roles of Pol B and Pol D in chromosomal DNA replication is great. Furthermore, on the basis of the results presented in chapter 5, even without a new reporter assay, the possibility of a previously unidentified TCR pathway in the archaea is an exciting prospect. The data presented in this thesis suggests that the proposed TCR pathway might occur in response to deaminated DNA, therefore a combination of *in vitro* analyses of archaeal RNA polymerase and additional genetic manipulation of *T. kodakarensis* should provide answers to the questions posed in chapter 5.

Finally, the study of uracil recognition by Pol D is likely to be aided by resolution of the whole structure of the enzyme by X-ray crystallography. Without such data, identification of the deaminated base recognition site of Pol D is likely to be extremely difficult. The preparation of mutants that retain polymerase activity but do not respond to uracil would be useful for both *in vitro* and *in vivo* studies. The possibility that deaminated base recognition relies on the formation of a tetrameric complex containing two small and two large subunits is also worthy of investigation. Size-exclusion chromatographic separation of DNA-bound complexes should provide an indication as to the conformation adopted by uracil-bound Pol D.

Chapter 8

References

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. (2002) Molecular biology of the cell. 4th edition. New York. Garland Science. ISBN: 0-8153-3218-1

Allers, T. and Mevarech, M. (2005) Archaeal genetics - the third way. *Nat. Rev. Genet.* **6:** 58–73

Allinson, S. L., Dianova, I. I. and Dianov, G. L. (2001) DNA polymerase ß is the major dRP lyase involved in repair of oxidative base lesions in DNA my mammalian cell extracts. *EMBO J.* **20:** 6919-6926

Aravind, L. and Koonin, E. V. (2000) The alpha/beta fold uracil DNA glycosylases: a common origin with diverse fates. *Genome Biol.* **1:** research0007

Astbury, W.T. and Bell, F.O. (1938) X-ray study of thymonucleic acid. *Nature* **141:** 747-748

Aves, S. J. (2009) DNA replication initiation. *Methods Mol. Biol.* **521:** 3-17

Baker, T. A. and Bell, S. P. (1998) Polymerases and the replisome: machines within machines. *Cell* **92:** 295-305

Bandwar, R. P. and Patel, S. S. (2001) Peculiar 2-aminoapurine fluorescence monitors the dynamics of open complex formation by bacteriophage T7 RNA polymerase. *J. Biol. Chem.* **276:** 14075–14082

Barry, E. R. and Bell, S. D. (2006) DNA replication in the archaea. *Microbiol. Mol. Biol. Rev.* **70:** 876-887

Bartlett, J. M. and Stirling, D. (2003) A short history of the polymerase chain reaction. *Methods Mol. Biol.* **226:** 3-6

Beechem, J. M., Otto, M. R., Bloom, L. B., Eritja, R., Reha-Krantz, L. J. and Goodman, M. F. (1998) Exonuclease-polymerase active site partitioning of primer-template DNA strands and equilibrium Mg^{2+} binding properties of bacteriophage T4 DNA polymerase. *Biochemistry* **37:** 10144–10155

Beese, L. S. and Steitz, T. A. (1991) Structural basis for the 3'-5' exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. *EMBO J.* **10(1):** 25-33

Bell, S. D., Dionne, I., Lao-Sirieix, S., Marsh, V. L., McGeoch, A. and Robinson, N. P. (2005) Archaeal DNA replication: a robust model for eukaryotes. *Geothermal Biology and Geochemsitry in Yellowstone National Park* 278-288

Berdis, A. J. (2008) DNA polymerases as therapeutic targets. *Biochemistry.* **47:** 8253-60

Berg, J. M., Tymoczko, J. L. and Stryer, L. (2006) Biochemistry. Sixth edition. W. H. Freeman. ISBN: 0-7167-8724-5

Berquist, B. R., DasSarma, P. and DasSarma, S. (2007) Essential and non-essential

DNA replication genes in the model halophilic archaeon Halobacterium sp. NRC-1. *BMC Genet.* **8:** article number 31

Bloom, L. M., Otto, M. R., Eritja, R., Reha-Krantz, L. J., Goodman, M. F. and Beechem, J. M. (1994) Pre-steady-state kinetic analysis of sequence-dependent nucleotide excision by the 3′-exonuclease activity of bacteriophage T4 DNA polymerase. *Biochemistry* **33:** 7576-7586

Bocquier, A. A., Liu, L., Cann, I. K. O., Komori, K., Kohda, D. and Ishino, Y. (2001) Archaeal primase: bridging the gap between RNA and DNA polymerases. *Curr. Biol.* **11:** 452-6

Braithwaite, D. K. and Ito, J. (1993) Compilation, alignment, and phylogenetic relationships of DNA polymerase. *Nucleic Acids Res.* **21:** 787-802

Brock, T. D., Brock, K. M., Belly, R. T. and Weiss, R. L. (1972) *Sulfolobus*: A new genus of sulfur-oxidizing bacteria living at low pH and high temperature. *Archives of Microbiology* **84:** 54-68

Brueckner, B., Kuck, D. and Lyko, F. (2007) DNA methyltransferase inhibitors for cancer therapy. *Cancer J.* **13:** 17-22

Burgers, P. M. (1998) Eukaryotic DNA polymerases in DNA replication and repair. *Chromosoma* **107:** 218-27

Burgers, P. M. and Eckstein, F. (1979) A study of the mechanism of DNA polymerase I from *Escherichia coli* with diastereomeric phosphorothioate analogs of deoxyadenosine triphosphate. *J. Biol. Chem.* **254:** 6889-93

Cann, I. K. O. and Ishino, Y. (1999) Archaeal DNA replication: identifying the pieces to solve a puzzle. *Genetics* **152:** 1249-1267

Cann, I. K., Komori, K., Toh, H., Kanai, S. and Ishino, Y. (1998) A heterodimeric DNA polymerase: evidence that members of the Euryarchaeota possess a distinct DNA polymerase. *Proc. Natl. Acad. Sci. USA* **95:** 14250-5

Capson, T. L., Paliska, J. A., Kaboord, B. F., Frey, M. W., Lively, C, Dahlberg, M. and Benkovic, S. J. (1992) Kinetic characterization of the polymerase and exonuclease activities of the gene 43 protein of bacteriophage T4. *Biochemsitry* **31:** 10984-94

Castle, P. E., Gutierrez, E. C., Leitch, S. V., Maus, C. E., McMillan, R. A., Nussbaumer, W. A., Vaughan, L. M., Wheeler, C. M., Gravitt, P. E. and Stiffman, M. (2011) An evaluation of a new DNA test for carcinogenic human papillomavirus. *J. Clin. Microbiol.* **49:** 3029-32

Castrec, B., Rouillon, C., Henneke, G., Flament, D., Querellou, J. and Raffin, J. P. (2009) Binding to PCNA in euryarchaeal DNA replication requires two PIP motifs for DNA polymerase D and one PIP motif for DNA polymerase B. *J. Mol. Biol.* **394:** 209-218

Chargaff, E., Lipschitz, R., Green C. and Hodes, M. E. (1951) The composition of

the deoxyribonucleic acid of salmon sperm. *J. Biol. Chem.* **192:** 223-230

Chen, H. and Shaw, B. R. (1993) Kinetics of bisulfite-induced cytosine deamination in single-stranded DNA. *Biochemistry* **32:** 3535-3539

Chien, A., Edgar, D. B. and Trela, J. M. (1976) Deoxyribonucleic acid polymerase from the extreme thermophile *Thermus aquaticus*. *J. Bacteriol.* **127:** 1550-1557

Cirino, P. C., Mayer, K. M. and Umeno, D. (2003) Generating mutant libraries using error-prone PCR. *Methods Mol. Biol.* **231:** 3-9

Cline, J., Braman, J. C. and Hogrefe, H. H. (1996) PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Res.* **24:** 3546-51

Cohen-Kupiec, R., Blank, C. and Leigh, J. A. (1997) Transcriptional regulation in archaea: *in vivo* demonstration of a repressor binding site in a methanogen. *Proc. Natl. Acad. Sci. USA* **94:** 1316–1320

Connolly, B. A. (2009) Recognition of deaminated by archaeal family-B DNA polymerases. *Biochem. Soc. Trans*. **37(Pt 1):** 65-8

Copeland, W. C. and Wang, T. S. (1993) Enzymatic characterization of the individual mammalian primase subunits reveals a biphasic mechanism for initiation of DNA replication. *J. Biol. Chem.* **268:** 26179-89

Cosman, M, de los Santos, C., Fiala, R., Hingerty, B. E., Ibanez, V., Margulis, L. A., Live, D., Geacintov, N. E. and Broyde, S. (1992) Solution conformation of the major adduct between the carcinogen (+)-anti-benzo[a]pyrene diol epoxide and DNA. *Proc. Natl. Acad. Sci. USA* **89:** 1914-1918

Cowan, J. A. (1998) Metal activation of enzymes in nucleic acid biochemistry. *Chem. Rev.* **98:** 1067–1087

Crumpacker, C. S. (1992) Mechanism of action of foscarnet against viral polymerases. *Am. J. Med.* **92(J):** 3S-7S

Dahm, R. (2008) Discovering DNA: Friedrich Meischer and the early years of nucleic acid research. *Human Genetics* **122:** 565-581

Dalhus, B., Arvai, A. S., Rosnes, I., Olsen, Ø. E., Backe, P. H., Alseth, I., Gao, H., Cao, W., Tainer, J. A. and Bjørås, M. (2009) Structures of endonuclease V with DNA reveal initiation of deaminated adenine repair. *Nat. Struct. Mol. Biol.* **16:** 138-43

Datta, K., Johnson, N. P., LiCata, V. J. and von Hippel, P. H. (2009) Local conformations and competitive binding affinities of single- and double-stranded primer-template DNA at the polymerization and editing active sites of DNA polymerases. *Journal of Biological Chemistry* **284:** 17180-93

Datta, K., Johnson, N. P. and von Hippel, P. H. (2010) DNA conformational changes at the primer-template junction regulate the fidelity of replication by DNA

polymerases. *Proc. Natl. Acad. Sci. USA* **107:** 17980-17985

Datta, R., Kharbanda, S. and Kufe, D. W. (1992) Transcriptional and posttranscriptional regulation of H1 histone gene expression by 1-beta-D-arabinofuranosylcytosine. *Mol. Pharmacol.* **41:** 64-8

Delarue, M, Poch, O., Tordo, N., Moras, D. and Argos, P. (1990) An attempt to unify the structure of DNA polymerases. *Protein Eng.* **3:** 461-7

Dizdaroglu, M., Karakaya, A., Jaruga, P., Slupphaug, G. and Krokan, H. E. (1996) Novel activities of human uracil DNA N-glycosylase for cytosine-derived products of oxidative DNA damage. *Nucleic Acids Res.* **24:** 418-22

Dominguez, O., Ruiz, J. F., Lain de Lera, T., Garcia-Diaz, M., González, M. A., Kirchhoff, T., Martinez-A, C., Bernad, A. and Blanco, L. (2000) DNA polymerase mu (Pol mu), homologous to TdT, could act as a DNA mutator in eukaryotic cells. *EMBO J.* **19:** 1731-42

Dong, X., Stothard, P., Forsythe, I. J. and Wishart, D. S. (2004) PlasMapper: a web server for drawing and auto-annotating plasmid maps. *Nucleic Acids Res.* **32(Web Server issue):** W660-4

Donlin, M. J., Patel, S. S. and Johnson, K. A. (1991) Kinetic partitioning between the exonuclease and polymerase sites in DNA error correction. *Biochemistry* **30:** 538-546

Dorazi, R., Götz, D., Munro, S., Bernander, R. and White, M. F. (2007) Equal rates of DNA photoproducts in transcribed and non-transcribed strands in *Sulfolobus solfitaricus*. *Mol. Microbiol.* **63:** 521-9

Duggin, I. G. and Bell, S. D. (2006) The chromosome replication machinery of the archaeon *Sulfolobus solfitaricus*. *J. Biol. Chem.* **281:** 15029-32

Edgell, D. R., Klenk, H. P. and Doolittle, W. F. (1997) Gene duplications in evolution of the archaeal family B DNA polymerases. *J. Bacteriol.* **179:** 2632-40

Eisen, J. and Hanawalt P. C. (1999) A phylogenomic study of DNA repair genes, proteins, and processes. *Mut. Res. DNA Repair* **435:** 171-213

Emptage, K. (2008) Role of replisome proteins in recognition of deaminated bases in archaea. Doctoral thesis. Newcastle University, UK.

Emptage, K., O'Neill, R., Solovyova, A. and Connolly, B. A. (2008) Interplay between DNA polymerase and proliferating cell nuclear antigen switches off base excision repair of uracil and hypoxanthine during replication in archaea. *J. Mol. Biol.* **383:** 762-771

Endoh, T., Kanai, T. and Imanaka, T. (2008) Effective approaches for the production of heterologous proteins using the *Thermococcus kodakarensis*-based translation system. *J. Biotechnol.* **133:** 177-82

Eritja, R., Horowitz, D. M., Walker, P. A., Ziehler-Martin, J. P., Boosalis, M. S.,

Goodman, M. F., Itakura, K. and Kaplan, B. E. (1986) Synthesis and properties of oligonucleotides containing 2′-deoxynebularine and 2′-deoxyxanthosine. *Nucleic Acids Res.* **14:** 8135−8153

Evans, S.J., Fogg, M. J., Mamone, A., Davis, M., Pearl, L. H. and Connolly, B. A. (2000) Improving dideoxynucleotide-triphosphate utilization by the hyper-thermophilic DNA polymerase from the archaeon *Pyrococcus furiosus*. *Nucleic Acids Res.* **28:** 1059-1066

Fagan, P. A., Fàbrega, C., Eritja, R., Goodman, M. F. and Wemmer, D. E. (1996) NMR study of the conformation of the 2-aminopurine:cytosine mismatch in DNA. *Biochemistry* **35:** 4026-4033

Feng, H., Klutz, A. M. and Cao, W. (2005) Active site plasticity of endonuclease V from *Salmonella typhimurium*. *Biochemistry* **44:** 675-83

Feng, Y., Zhuang, W. and Prohofsky, E. W. (1990) Calculation of temperature dependence of interbase breathing motion of a guanine-cytosine DNA double helix with adenine-thymine insert. *Phys. Rev. A.* **43:** 1049-1053

Fiala, G., Stetter, K. O. (1986). *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaebacteria growing optimally at 100°C. *Archives of Microbiology* **145:** 56–6

Firbank, S. J., Wardle, J., Heslop, P., Lewis, R. J. and Connolly, B. A. (2008) Uracil recognition in archaeal DNA polymerases captured by X-ray crystallography. *J. Mol. Biol.* **381:** 529-39

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A., Merrick, J. M. *et al.* (1995) Whole-genome random sequencing and assembly of *Haemophilus influenziae* Rd. *Science* **269:** 496-512

Fogg, M. J., Pearl, L. H. and Connolly, B. A. (2002) Structural basis for uracil recognition by archaeal family B DNA polymerases. *Nat. Struct. Biol.* **9:** 922-7

Franklin, M. C., Wang, J. and Steitz, T. A. (2001) Structure of the replicating complex of a Pol α family DNA polymerase. *Cell* **105:** 657-667

Franklin, R. E. (1953) Molecular configuration in sodium thymonucleate. *Nature* **171:** 740-741

Frederico, L. A., Kunkel, T. A. and Shaw, B. R. (1990) A sensitive genetic assay for the detection of cytosine deamination: determination of rate constants and the activation energy. *Biochemistry*. **29:** 2532-7

Freed, E. O. and Martin, M. A. (2001) HIVs and their replication. In Knipe, D. and Howley, P. (eds) Fields Virology. Vol. 2. Philadelphia. Lippincott Williams & Wilkins.

Freemont, P. S., Friedman, J. M., Beese, L. S., Sanderson, M. R. and Steitz, T. A. (1988) Cocrystal structure of an editing complex of Klenow fragment with DNA.

Proc. Natl. Acad. Sci. USA **85:** 8924-8928

Friedrich-Heineken, E. and Hübscher, U. (2004) The Fen1 extrahelical 3'-flap pocket is conserved from archaea to human and regulates DNA substrate specificity. *Nucleic Acids Res.* **32:** 2520-2528

Friedberg, E. C., Wagner, R. and Radman, M. (2002) Specialized polymerases, cellular survival, and the genesis of mutations. *Science* **296:** 1627-30

Friedman, J. I. and Stivers, J. T. (2010) Detection of damaged bases by DNA glycosylase enzymes. *Biochemistry* **49:** 4957-4967

Gan, G. N., Wittschieben, J. P., Wittschieben, B. Ø. and Wood, R. D. (2008) DNA polymerase zeta (pol zeta) in higher eukaryotes. *Cell Res.* **18:** 174-83

Garg, P., Stith, C. M., Sabouri, N., Johansson, E. and Burgers, P. M. J. (2004) Idling by DNA polymerase delta maintains a ligatable nick during lagging-strand DNA replication. *Genes Dev.* **18:** 2764–2773

Garvik, B., Carson, M. and Hartwell, L. (1995) Single –stranded DNA arising at telomeres in *cdc13* mutants may constitute a specific signal for the *RAD9* checkpoint. *Mol. Cell Biol.* **15:** 6128-38

Garg, P. and Burgers, P. M. (2005) DNA polymerases that propagate the eukaryotic DNA replication fork. *Crit. Rev. Biochem. Mol. Biol.* **40:** 115-28

Gildea, B. and McLaughlin, L. W. (1989) The synthesis of 2-pyrimidinone nucleosides and their incorporation into oligodeoxynucleotides. *Nucleic Acids Res.* **17:** 2261-81

Gill, S., O'Neill, R., Lewis, R. J. and Connolly, B. A. (2007) Interaction of the family-B DNA polymerase from the archaeon *Pyrococcus furiosus* with deaminated bases. *J. Mol. Biol.* **372:** 855-63

Goodman, M. F. (2002) Error-prone repair DNA polymerases in prokaryotes and eukaryotes. *Annu. Rev. Biochem.* **71:** 17-50

Greagg, M. A., Fogg, M. J., Panayatou, G., Evans, S. J., Connolly, B. A. and Pearl, L. H. (1999) A read-ahead function in archaeal DNA polymerases detects promutagenic template-strand uracil. *Proc. Natl. Acad. Sci. USA* **96:** 9045-9050

Gueguen, Y., Roland, J. L., Lecompte, O., Azam, P., Le Romancer, G., Flament, D., Raffin, J. P. and Dietrich, J. (2001) Characterization of two DNA polymerases from the hyperthermophilic euryarchaeon *Pyrococcus abyssi*. *Eur. J. Biochem.* **268:** 5961-9

Guest, C. R., Hochstrasser, R. A., Sowers, L. C. and Millar, D. P. (1991) Dynamics of mismatched base pairs. *Biochemistry* **30:** 3271-3279

Gutman, P. D. and Minton, K. W. (1993) Conserved sites in the 5'-3' exonuclease domain of *Escherichia coli* DNA polymerase. *Nucleic Acids Res.* **21:** 4406-4407

Hagen, L., Peña-Diaz, J., Kavli, B., Otterlei, M., Slupphaug, G. and Krokan, H. E. (2006) Genomic uracil and human disease. *Exp. Cell Res.* **312:** 2666-72

Hanawalt, P. C. (1994) Transcription-coupled repair and human disease. *Science* **266:** 1957-8

Hariharan, C. and Reha-Krantz, L. J. (2005) Using 2-aminopurine fluorescence to detect bacteriophage T4 DNA polymerase-DNA complexes that are important for primer extension and proofreading reactions. *Biochemistry* **44:** 15674-15684

Henneke, G., Flament, D., Hübscher, U., Querellou, J. and Raffin, J. (2005) The hyperthermophilic euryarchaeota *Pyrococcus abyssi* likely requires the two DNA polymerases D and B for DNA replication. *Journal of Molecular Biology* **350:** 53-64

Hinks, J. A. Evans, M. C., De Miguel, Y., Sartori, A. A., Jiricny, J. and Pearl, L. H. (2002) An iron-sulfur cluster in the family 4 uracil-DNA glycosylases. *J. Biol. Chem.* **277:** 16936-16940

Hogg, M., Aller, P., Konigsberg, W., Wallace, S. S., and Doublie, S. (2007) Structural and biochemical investigation of the role in proofreading of a β hairpin loop found in the exonuclease domain of a replicative DNA polymerase of the B family. *J. Biol. Chem.* **282:** 1432–1444

Hogg, M., Wallace, S. S. and Doublié, S. (2004) Crystallographic snapshots of a replicative DNA polymerase encountering an abasic site. *EMBO J.* **23:** 1483-93

Horton, R. M., Hunt, H. D., Ho., S. N., Pullen, J. K. and Pease, L. R. (1989) Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene* **77:** 61-68

Huang, P., Chubb, S. and Plunkett, W. (1990) Termination of DNA synthesis by 9-beta-D-arabinofuranosyl-2-fluoroadenine. A mechanism for cytotoxicity. *J. Biol. Chem.* **265:** 16617-25

Idris, H. T., Al-Assar, O. and Wilson, S. H. (2002) DNA polymerase β. *Int. J. Biochem. Cell Biol.* **34:** 321-4

Indiani, C. and O'Donnell, M. (2006) The replication clamp-loading machine at work in the three domains of life. *Nat. Rev. Mol. Cell Biol.* **7:** 751-61

Ishino, Y. and Ishino, S. (2001) DNA polymerases from euryarchaeota. *Methods in Enzymology* **334:** 249-260

Ishino Y., Komori K., Cann I.K., Koga Y. (1998) A novel DNA polymerase family found in archaea. *J. Bacteriol.* **180:** 2232-2236

Ito, J. and Braithwaite, D. K. (1991) Compilation and alignment of DNA polymerase sequences. *Nucleic Acids Res.* **19:** 4045-57

Jacobi, C. A., Roggenkamp, A., Rakin, A., Zumbihi, R., Leitritz, L. and Heesemann, J. (1998) *In vitro* and *in vivo* expression studies of *yopE* from *Yersinia enterocolitica* using the *gfp* reporter gene. *Mol. Microbiol.* **30:** 865–882

Jean, J. M. and Hall, K. B. (2001) 2-aminopurine fluorescence quenching and lifetimes: role of base stacking. *Proc. Natl. Acad. Sci. USA* **98:** 37-41

Johnson, K. A. (1993) Conformational coupling in DNA polymerase fidelity. *Annu. Rev. Biochem.* **62:** 685-713

Johnson, R. E., Prakash, S. and Prakash, L. (2000) The human DINB1 gene encodes the DNA polymerase Pol θ. *Proc. Natl. Acad. Sci. USA* **97:** 3838–3843

Johnson, S. J. and Beese, L. S. (2004) Structures of mismatch replication errors observed in a DNA polymerase. *Cell* **116:** 803-816

Joyce C. M. (1989) How DNA travels between the separate polymerase and 3′–5′ exonuclease sites of DNA polymerase I (Klenow fragment). *J. Biol. Chem.* **264:** 10858–10866

Joyce, C. M. (2010) Techniques used to study the DNA polymerase reaction pathway. *Biochimia Biophysica Acta* **1804:** 1032–1040

Joyce, C. M. and Steitz, T. A. (1994) Function and structure relationships in DNA polymerases. *Annu. Rev. Biochem.* **63:** 777-822

Kamtekar, S., Berman, A. J., Wang, J., Lázaro, J. M., de Vega, M., Blanco, L., Salas, M. and Steitz, T. A. (2004) Insights into strand displacement and processivity from the crystal structure of the protein-primed DNA polymerase of bacteriophage phi29. *Mol. Cell* **16:** 609-18

Kelman, Z. and O'Donnell, M. (1995) DNA polymerase III holoenzyme: structure and function of a chromosomal replicating machine. *Annu. Rev. Biochem.* **64:** 171-200

Khare, V. and Eckert, K. A. (2001) The 3'-5' exonuclease of T4 DNA polymerase removes premutagenic alkyl mispairs and contributes to futile cycling at 0^6 methylguanine lesions. *J. Biol. Chem.* **276:** 24286–24292

Khare, V. and Eckert, K. A. (2002) The proofreading 3'–5' exonuclease activity of DNA polymerases: a kinetic barrier to translesion DNA synthesis. *Mutation Res.* **510:** 45–54

Killelea, T., Ghosh, S., Tan, S. S., Heslop, P., Firbank, S. J., Kool, E. T. and Connolly, B. A. (2010) Probing the interaction of archaeal DNA polymerases with deaminated bases using X-ray crystallography and non-hydrogen bonding isosteric base analogues. *Biochemistry* **49:** 5772-81

Klein, D. (2002) Quantification using real-time PCR technology: applications and limitations. *Trends Mol Med.* **8:** 257-60

Kornberg, A. I. R., Lehman and Simms, E. S. (1956) Polydesoxyribonucleotide synthesis by enzymes from *Escherichia coli*. *Fed. Proc.* **15:** 291-292

Kunkel, T. A. (1988) Exonucleolytic proofreading. *Cell* **53:** 837-840

Kunkel, T. A. and Bebenek, K. (2000) DNA replication fidelity. *Annu. Rev. Biochem.* **69:** 497-529

Law, S. M., Eritja, R., Goodma, M. F. and Breslauer, K. J. (1996) Spectroscopic and calorimetric characterizations of DNA duplexes containing 2-aminopurine. *Biochemistry* **35:** 12329-12337

Lee, I. and Berdis, A. (2006) Fluorescent analysis of translesion DNA synthesis by using a novel, non-natural nucleotide analogue. *Chembiochem* **7:** 1990-1997

Lehman, I. R., Bessman, M. J., Simms, E. S. and Kornberg, A. (1958) Enzymatic synthesis of deoxyribonucleic acid. *J. Bio. Chem.* **233:** 163-170

Lieber, M. R. (1997) The FEN-1 family of structure-specific nucleases in eukaryotic DNA replication recombination and repair. *Bioessays* **19:** 233-40

Lin, L. J., Yoshinaga, A., Lin, Y., Guzman, C., Chen, Y. H., Mei, S., Lagunas, A. M., Koike, S., Iwai, S., Spies, M. A., Nair, S. K., Mackie, R. I., Ishino, Y and Cann, I. K. (2010) Molecular analyses of an unusual translesion DNA polymerase from *Methanosarcina acetivorans* C2A. *J. Mol. Biol.* **397:** 13-30

Lin, T. C., Karam, G., and Konigsberg, W. H. (1994). Isolation, characterization, and kinetic properties of truncated forms of T4 DNA polymerase that exhibit 3'-5' exonuclease activity. *J. Biol. Chem.* **269:** 19286–19294

Lin, X., Liu, J., Maley, F. and Chu, E. (2003) Role of cysteine amino acid residues on the RNA binding activity of human thymidylate synthase. *Nucleic Acids Res*. **31:** 4882-4887

Lindahl, T., Ljungquist, S., Siegert, W., Nyberg, B. and Sperens, B. (1977) DNA N-glycosidases: properties of uracil-DNA glycosidase from *Escherichia coli*. *J. Biol. Chem.* **252:** 3286-94

Liu, Y., Kao, H.-I. and Bambara, R. A. (2004) Flap endonuclease 1: a central component of DNA metabolism. *Annu. Rev. Biochem.* **73:** 589-615

Liu, L., Komori, K., Ishino, S., Bocquier, A. A., Cann, I. K. O., Kohda, D. and Ishino, Y. (2001) The archaeal DNA primase: biochemical characterization of the p41-p46 complex from *Pyrococcus furiosus*. *J. Biol. Chem.* **276:** 45484-90

Liu, C. and Martin, C. T. (2002) Promoter clearance by T7 RNA polymerase. Initial bubble collapse and transcript dissociation monitored by base analog fluorescence. *J. Biol. Chem.* **277:** 2725–2731

Longo, M. C., Berninger, M. S. and Hartley, J. L. (1990) Use of uracil DNA glycosylase to control carry-over contamination in polymerase chain reactions. *Gene* **93:** 125-8

Lundberg, K. S., Shoemaker, D. D., Adams, M. W., Short, J. M., Sorge, J. A. and Mathur, E. J. (1991) High-fidelity amplification using a thermostable DNA polymerase isolated from *Pyrococcus furiosus*. *Gene* **108:** 1-6

Marquez, L. A. and Reha-Krantz, L. J. (1996) Using 2-aminopurine fluorescence and mutational analysis to demonstrate an active role of bacteriophage T4 DNA polymerase in strand separation required for 3'-5' exonucleolytic activity. *J Biol Chem.* **271:** 28903-11

Mason, M., Schuller, A. and Skordalakes, E. (2011) Telomerase structure function. *Curr. Opin. Struct. Biol.* **21:** 92-100

Matsui, E., Nishio, M, Yokoyama, H., Harata, K., Darnis, S. and Matsui, I (2003) Distinct domain functions regulating *de novo* synthesis of thermostable DNA primase from hyperthermophile *Pyrococcus horikoshii*. *Biochemistry* **42:** 14968-14976

Matsui, I., Urushibata, Y., Shen, Y., Matsui, E. and Yokoyama, H. (2011) Novel structure of an terminal domain that is crucial for the dimeric assembly and DNA-binding of an archaeal DNA polymerase D large subunit from *Pyrococcus horikoshii*. *FEBS Lett.* **585:** 452-458

Matsunaga, F., Norais, C., Forterre, P. and Myllykallio, H. (2003) Identification of short "eukaryotic" Okazaki fragments synthesised from a prokaryotic replication origin. *EMBO Rep.* **4:** 154-8

Mellon, I. and Hanawalt, P. C. (1989) Induction of the *Escherichia coli* lactose operon selectively increases repair of its transcribed DNA strand. *Nature* **342:** 95-8

Mendel, G. (1865) Experiments in plant hybridisation. *Meetings of the Brünn Natural History Society* 3-47

Meselson, M. and Stahl, F. W. (1958) The replication of DNA in *E. coli*. *Proc. Natl. Acad. Sci. USA* **44:** 671-682

Meyer, R. P. and Laine, P. S. (1990) The single-stranded binding protein of *Escherichia coli*. *Microbiol Rev.* **54:** 342-380

Mi, R., Abole, A. K., Cao, W. (2011) Dissecting the endonuclease and exonuclease activities in endonuclease V from *Thermotoga maritime*. *Nucleic Acids Res.* **39:** 536-44

Miller, H. and Grollman, A. P. (1997) Kinetics of DNA polymerase I (Klenow fragment exo-) activity on damaged DNA templates: effect of proximal and distal template damage on DNA synthesis. *Biochemistry* **36:** 15336-42

Miller, J. H. (1972) Experiments in molecular genetics (Cold Spring Harbor Lab., Cold Spring Harbor, NY)

Miura, S. and Izuta, S. (2004) DNA polymerases as targets of anticancer nucleosides. *Curr. Drug Targets.* **5:** 191-5

Moldovan, G. L., Pfander, B. and Jentsch, S. (2007) PCNA, the maestro of the replication fork. *Cell.* **129:** 665-79

Neely, R. K., Daujotyte, D., Grazulis, S., Magennis, S. W., Dryden, D. T., Klimasauskas, S. and Jones, A. C. (2005) Time-resolved fluorescence of 2-

aminopurine as a probe of base flipping in M.HhaI-DNA complexes. *Nucleic Acids Res.* **33:** 6953-60

Neely R. K. , Magennis S. W. , Dryden D. T. F. , Jones A. C. (2004) Evidence of tautomerism in 2-aminopurine from fluorescence lifetime measurements. *J. Phys. Chem. B.* **108:** 17606–17610

Ng, L., Weiss, S. J., and Fisher, P. A. (1989) Recognition and binding of template-primers containing defined abasic sites by *Drosophila* DNA polymerase alpha holoenzyme. *J. Biol. Chem.* **264:** 13018-13023

Nick McElhinny, S. A., Gordenin, D. A., Stith, C. M., Burgers, P. M. and Kunkel, T. A. (2008) Division of labor at the eukaryotic replication form. *Mol. Cell.* **30:** 137-44

Nilsen, H., Haushalter, K. A., Robins, P., Barnes, D. E., Verdine, G. L. and Lindahl, T. (2001) Excision of deaminated cytosine from the vertebrate genome: role of the SMUG1 uracil-DNA glycosylase. *EMBO J.* **20:** 4278-86

Nordlund, T. M., Andersson, S., Nilsson, L. and Rigler, R. (1989) Structure and dynamics of a fluorescent DNA oligomer containing the *Eco*RI recognition sequence: fluorescence, molecular dynamics and NMR studies. *Biochemistry* **28:** 9095-9103

O'Donnell, M. (2006) Replisome architecture and dynamics in *Escherichia coli*. *J Biol Chem.* **281:** 10653-6

Oakley, G. G. and Patrick, S. M. (2010) Replication protein A: directing traffic at the intersection of replication and repair. *Front. Biosci.* **15:** 883-900

Oberg, B. (1989) Antiviral effects of phosphonoformate (PFA, Foscarnet sodium). *Pharmacol. Ther.* **40:** 213-85

Őberg, B. (2006) Rational design of polymerase inhibitors as antiviral drugs. *Antiviral Research* **71**: 90-95

Ogawa, T. and Okazaki, T. (1984) Function of RNaseH in DNA replication revealed by RNase H defective mutants of *Escherichia coli*. *Mol. Cell. Genet.* **193:** 231-7

Ohashi, E., Bebenek, K., Matsuda, T., Feaver, W. J., Gerlach, V. L., Friedberg, E. C., Ohmori, H. and Kunkel, T. A. (2000) Fidelity and processivity of DNA synthesis by DNA polymerase κ, the product of the human DINB1 gene. *J. Biol. Chem.* **275:** 39678–39684

Olerup, O. and Zetterquist, H. (1992) HLA-DR typing by PCR amplification with sequence-specific primers (PCR-SSP) in 2 hours: an alternative to serological DR typing in clinical practice including donor-recipient matching in cadaveric transplantation. *Tissue antigens.* **39:** 225-35

Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G. and Steitz, T. A. (1985) Structure of the large fragment of *Escherichia coli* DNA polymerase I complexed with dTMP. *Nature* **313:** 762-766

Otterlei, M., Kavli, B., Standal, R., Skjelbred, C., Bharati, S. and Krokan, H. E. (2000) Repair of chromosomal abasic sites *in vivo* involves at least three different repair pathways. *EMBO J.* **19:** 5542-51

Otvos, J. D. and Armitage, I. M. (1980) Structure of the metal clusters in rabbit liver metallothionein. *Proc. Natl. Acad. Sci. USA.* **77:** 7094–7098

Passerini, A., Punta, M., Ceroni, A., Rost, B. and Frasconi, P. (2006) Identifying cysteines and histidines in transition-metal-binding sites using support vector machines and neural networks. *Proteins* **65:** 305-316

Parker, W. B., Allan, P. W., Shaddix, S. C., Rose, L. M., Speegle, H. F., Gillespie, G. Y., Bennett, L. L. Jr. (1998) Metabolism and metabolic actions of 6-methylpurine and 2-fluoroadenine in human cells. *Biochem. Pharmacol.* **55:** 1673-1681

Pelletier, H., Sawaya, M. R., Kumar, A., Wilson, S. H. and Kraut, J. (1994) Structures of the ternary complexes of rat DNA polymerase beta, a DNA template-primer, and ddCTP. *Science* **264:** 1891-903

Plato (c. 375 BC) Myth of the metals. *The Republic* **Book III**

Plosky, B. S. and Woodgate, R. (2004) Switching from high-fidelity polymerases to low-fidelity lesion-bypass polymerases. *Curr. Opin. Genet. Dev.* **14:** 113-9

Porecha, R. H. and Stivers, J. T. (2008) Uracil DNA glycosylase uses DNA hopping and short-range sliding to trap extrahelical uracils. *Proc. Natl. Acad. Sci. USA* **105:** 10791-6

Prasad, R., Longley, M. J., Sharief, F. S., Hou, E. W., Copeland, W. C. and Wilson, S. H. (2009) Human DNA polymerase θ possesses 5′-dRP lyase activity and functions in single-nucleotide base excision repair *in vitro*. *Nucleic Acids Res.* **37:** 1868-77

Pribnow, D. (1975) Nucleotide sequence if an RNA polymerase binding site at an early T7 promoter. *Proc. Natl. Acad. Sci. USA* **72:** 784-788

Prober, J. M., Trainor, G. L., Dam, R. J., Hobbs, F. W., Robertson, C. W., Zagursky, R. J., Cocuzza, A. J., Jensen, M. A. and Baumeister, K. (1987) A system for rapid DNA sequencing with fluorescent chain terminating dideoxynucleotides. *Science.* **238:** 336-341

Pursell, Z. F., Isoz, I., Lundström, E. B., Johansson, E. and Kunkel, T. A. (2007) Yeast DNA polymerase epsilon participates in leading strand DNA replication. *Science*. **317:** 127-30

Purves, W. K., Orians, G. H., Craig Heller, H. and Sadava, D. (2001) Life: the science of biology. 5th edition. W. H. Freeman & Co Ltd. ISBN: 0-7167-3325-0

Rajilić-Stojanović, M., Smidt, H. and de Vos, W. M. (2007) Diversity of the human gastrointestinal tract microbiota revisted. *Environ Microbiol.* 9**:** 2125-36

Ramadan, K., Shevelev, I. and Hübscher, U. (2004) The DNA polymerase-X

family: controllers of DNA quality? *Nat. Rev. Mol. Cell Biol.* **5:** 1038-43

Rattray, A. J. and Strathern, J. N. (2003) Error-prone DNA polymerases: when making a mistake is the only way to get ahead. *Annu. Rev. Gene.* **37:** 31-66

Reha-Krantz, L. J. (2010) DNA polymerase proofreading: Multiple roles maintain genome stability. *Biochim. Biophys. Acta.* **1804:** 1049–1063

Reid, S. L., Parry, D., Liu, H. H. and Connolly, B. A. (2001) Binding and recognition of GATATC target sequences by the EcoRV restriction endonuclease: a study using fluorescent oligonucleotides and fluorescence polarization. *Biochemistry* **40:** 2484-94

Rist, M. J. and Marino, J. P. (2002) Fluorescent nucleotide base analogs as probes of nucleic acid structure, dynamics and interactions. *Curr. Org. Chem.* **6:** 775–793

Robinson, N. P., Dionne, I., Lundgren, M., Marsh, V. L., Bernander, R. and Bell, S. D. (2004) Identification of two origins of replication in the single chromosome of the archaeon *Sulfolobus solfitaricus*. *Cell* **116:** 25-38

Rogozin, I. B., Makarova, K. S., Pavlov, Y. I. and Koonin, E. V. (2008) A highly conserved family of inactivated archaeal B family DNA polymerases. *Biol. Direct.* **3:** 32

Romano, V., Napoli, A., Salerno, V., Valenti, A., Rossi, M. and Ciaramella, M. (2007) Lack of strand-specific repair of UV-induced DNA lesions in three genes of the archaeon *Sulfolobus solfataricus*. *J. Mol. Biol.* **365:** 921-9

Rothwell, P. J. and Waksman, G. (2005) Structure and mechanism of DNA polymerases. *Adv. Protein Chem.* **71:** 401-40

Rouillon, C., Henneke, G., Flament, D., Querellou, J., Raffin, J-P. (2007) DNA polymerase switching on homotrimeric PCNA at the replication fork of the euryarchaea *Pyrococcus abyssi*. *J. Mol. Biol.* **369:** 343-355

Rudolf, J., Makrantoni, V., Ingledew, W. J., Stark, M. J. and White, M. F. (2006) The DNA repair helicases XPD and FancJ have essential iron-sulfur domains. *Mol. Cell* **23:** 801-808

Russell, H. J., Richardson, T. T., Emptage, K. and Connolly, B. A. (2009) The 3'-5' exonuclease of archaeal family-B DNA polymerase hinders copying of template strand deaminated bases. *Nucleic Acids Res.* **37:** 7603-11

Saiki, R. K., Gelfand, D. K., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. and Erlich, H. A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239:** 487-491

Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A. and Arnheimm, N. (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anaemia. *Science* **230:** 1350-1354

Saladin, K. S. (1998) Anatomy and Physiology: the unity of form and function. Boston. McGraw-Hill.

Sanger, F., Nicklen, S. and Coulson, A. R. (1977) DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* **74:** 5463-7

Santangelo T. J., Cubonová L., Reeve J. N. (2010) *Thermococcus kodakarensis* genetics: TK1827-encoded β-glycosidase, new positive-selection protocol, and targeted and repetitive deletion technology. *Appl. Environ. Microbiol.* **76:** 1044–1052

Sato, T., Fukui, T., Atomi, H. & Imanaka, T. (2003). Targeted gene disruption by homologous recombination in the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. *J. Bacteriol.* **185:** 210–220

Sato, T., Imanaka, H., Rashid, N., Fukui, T., Atomi, H. and Imanaka, T. (2004) Genetic evidence identifying the true gluconeogenic fructose-1,6-bisphosphatase in *Thermococcus kodakaraensis* and other hyperthermophiles. *J. Bacteriol.* **186:** 5799-5807

Savery, N. J. (2011) Prioritizing the repair of DNA damage that is encountered by RNA polymerase. *Transcription.* **2:** 168-172

Sawai H., Nagashima, J., Kuwahara, M., Kitagata, R., Tamura, T. and Matsui, I. (2007) Differences in substrate specificity of C(5)-Substituted or C(5)-unsubstituted pyrimidine nucleotides by DNA polymerases from thermophilic bacteria, archaea, and phages. *Chemistry and Biodiversity* **4:** 1979-1995

Sawaya, M. R., Pelletier, H., Kumar, A., Wilson, S. H. and Kraut, J. (1994) Crystal structure of rat DNA polymerase beta: evidence for a common polymerase mechanism. *Science* **264:** 1930-5

Sayers, J. R. (1994) Computer aided identification of a potential 5'-3' exonuclease gene encoded by *Escherichia coli*. *J. Theor. Biol.* **170:** 415-21

Schouten, K. A. and Weiss, B. (1999) Endonuclease V protects *Escherichia coli* against specific mutations caused by nitrous acid. *Mutat. Res.* **435:** 245-54

Schroeder, G. K. and Wolfenden, R. (2007) Rates of spontaneous disintegration of DNA and the rate enhancements produced by DNA glycosylases and deaminases. *Biochemistry* **46:** 13638-47

Schwartz, H., Shavitt, O. and Livneh, Z. (1988) The role of exonucleolytic processing and polymerase-DNA association in vitro: significance for SOS targeted mutagenesis. *J. Biol. Chem.* **263:** 18277–18285

Selby, C. P. and Sancar, A. (1993) Molecular mechanism of transcription-repair coupling. *Science* **260:** 53-8

Service, R. F. (2001) PCR. Roche dealt a setback on European Taq patent. *Science* **292:** 1815

Shamoo, Y. and Steitz, T. A. (1999) Building a replisome from interacting pieces:

sliding clamp complexed to a peptide from DNA polymerase and a polymerase editing complex. *Cell* **99:** 155-66

Shapiro, R. and Pohl, S.H. (1968) The reaction of ribonucleosides with nitrous acid. Side products and kinetics. *Biochemistry* **7:** 448–455

She, Q., Singh, R. M., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M. J., Chan-Weiher, C. C., Clausen I. G., Curtis B. A., De Moors A. *et al.* (2002) The complete genome of the crenarchaeon *Sulfolobus solfitaricus* P2. *Proc. Natl. Acad. Sci. USA* **98:** 7835-7840

Shen, Y., Tang, X. F. and Matsui, I. (2003) Subunit interaction and regulation of activity through terminal domains of the family D DNA polymerase from *Pyrococcus korikoshii*. *J. Biol. Chem.* **278:** 21247-57

Shuttleworth, G., Fogg, M. J., Kurpiewski, M. R., Jen-Jacobson, L. and Connolly, B. A. (2004) Recognition of the pro-mutagenic base uracil by family B DNA polymerases from archaea. *J. Mol. Biol.* **337:** 621-34

Sowers, L. C., Boulard, Y. and Fazakarley, G. V. (2000) Multiple structures for the 2-aminopurine-cytosine mispair. *Biochemistry* **39:** 7613-7620

Sowers L. C., Fazakerley G. V., Eritja R., Kaplan B. E., Goodman M. F. (1986) Base-pairing and mutagenesis - observation of a protonated base pair between 2-aminopurine and cytosine in an oligonucleotide by proton NMR. *Proc. Natl Acad. Sci. USA* **83:** 5434-5438

Steitz, T. A., Smerdon, S. J., Jäger, J., Joyce, C. M. (1994) A unified polymerase mechanism for nonhomologous DNA and RNA polymerases. *Science* **266:** 2022-5

Steitz, T. A. (1999) DNA polymerases: Structural diversity and common mechanisms. *J. Biol. Chem.* **274:** 17395-8

Stetter, K. O. (1982) Ultrathin mycelia-forming organisms from submarine volcanic areas having an optimum growth temperature of 105°C. *Nature* **300:** 258–260

Stivers, J. T. (1998) 2-aminopurine fluorescence studies of base stacking interactions at abasic sites in DNA: metal-ion and base sequence effects. *Nucleic Acids Res.* **26:** 3837–3844

Su, T. J., Connolly, B. A., Darlington, C., Mallin, R., Dryden, D. T. F. (2004) Unusual 2-aminopurine fluorescence from a complex of DNA and the EcoKI methyltransferases. *Nucleic Acids Res.* **32:** 2223–2230

Subach, O. M., Khoroshaev, A. V., Gerasimov, D. N., Baskunov, V. B., Shchyolkina, A. K. and Gromova, E. S. (2004) 2-Pyrimidinone as a probe for studying the EcoRII DNA methyltransferase-substrate interaction. *Eur. J. Biochem.* **271:** 2391-9

Subuddhi, U., Hogg, M. and Reha-Krantz, L. J. (2008) Use of 2-aminopurine fluorescence to study the role of the beta hairpin in the proofreading pathway catalyzed by the phage T4 and RB69 DNA polymerase. *Biochemistry* **47:** 6130-7

Svejstrup, J. Q. (2002) Mechanisms of transcription-coupled DNA repair. *Nat rev Mol Cell Biol.* **3:** 21-29

Tahirov, T. H., Makarova, K. S., Rogozin, I. B., Pavlov, Y. I. and Koonin, E. V. (2009) Evolution of DNA polymerases: an inactivated polymerase-exonuclease module in Pol epsilon and a chimeric origin of eukaryotic polymerases from two classes of archaeal ancestors. *Biol. Direct.* **4:** 11

Takasawa, K., Masutani, C., Hanaoka, F. and Iwai, S. (2004) Chemical synthesis and translesion replication of a cis-syn cyclobutane thymine-uracil dimer. *Nucleic Acids Res.* **32:** 1738-45

Tang, M., Pham, P., Shen, X., Taylor, J. S., O'Donnell, M., Woodgate, R. and Goodman, M. F. (2000) Roles of *E. coli* DNA polymerases IV and V in lesion-targeted and untargeted SOS mutagenesis. *Nature* **404:** 1014–1018

Tang, X. F., Shen, Y., Matsui, E. and Matsui, I. (2004) Domain topology of the DNA polymerase D complex from a hyperthermophilic archaeon *Pyrococcus horikoshii*. *Biochemistry* **43:** 11818-27

Trzenecka, A., Pzochocka, D., and Bebenek, A. (2009) Different behaviors in vivo of mutations in the β hairpin loop of the DNA polymerases of the closely related phages T4 and RB69. *J. Mol. Biol*. **289:** 797–807

Tye, B. K. (2000) Insights into DNA replication from the third domain of life. *Proc. Natl. Acad. Sci. USA* **97:** 2399-2401

Uemori, T., Sato, Y., Kato, I., Doi, H. and Ishino, Y. (1997) A novel DNA polymerase in the hyperthermophilic archaeon, *Pyrococcus furiosus*: gene cloning, expression, and characterization. *Genes Cells.* **2:** 499-512

Vaisman, A., Taksawa, K., Iwai, S. and Woodgate, R. (2006) DNA polymerase iota-dependent translesion synthesis replication of uracil containing cyclobutane pyrimidine dimers. *DNA Repair (Amst.).* **5:** 210-8

Wang, J., Sattar, A. K., Wang, C. C., Karam, J. D., Konigsberg, W. H. and Steitz, T. A. (1997) Crystal structure of a Pol alpha family replication DNA polymerase from bacteriophage RB69. *Cell* **89:** 1087-99

Wang, Y., Prosen, D. E., Mei, L., Sullivan, J.C., Finney, M. and Vander Horn, P. B. (2004) A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance *in vitro*. *Nucleic Acids Res.* **32:** 1197-1207

Wardle, J., Burgers, P. M., Cann, I. K., Darley, K., Heslop, P., Johansson, E., Lin, L. J., McGlynn, P., Sanvoisin, J., Stith, C. M. and Connolly, B. A. (2008) Uracil recognition by replicative polymerases is limited to the archaea, not occurring with bacteria and eukarya. *Nucleic Acids Res.* **36:** 705-11

Watson, J. D. and Crick, F. H. C. (1953) A structure for deoxyribose nucleic acid. *Nature* **171:** 737-738

Weiss, B. (2008) Removal of deoxyinosine from the *Escherichia coli* chromosome as

studied by oligonucleotide transformation. *DNA Repair (Amst.).* **17:** 205-12

Whittaker, R. H. (1969) New concepts of kingdoms or organisms. Evolutionary relations are better represented by new classifications than by the traditional two kingdoms. *Science* **163:** 150-160

Willard, J. M., Lee, D. A. and Holland, M. M. (1998) Recovery of DNA for PCR amplification from blood and forensic samples using a chelating resin. *Methods Mol. Biol.* **98:** 9-18

Woese, C. R., Kandler, O. and Wheelis, M. L. (1990) Towards a natural system of organism: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. USA* **87:** 4576-9

Wobbe, C. R. and Struhl, K. (1990) Yeast and human TATA-binding proteins have nearly identical DNA sequence requirements for transcription *in vitro*. *Mol. Cell Biol.* **10:** 3859-3867

Yamasaki, K., Urushibata, Y., Yamasaki, T., Arisaka, F. and Matsui, I. (2010) Solution structure of the N-terminal domain of the archaeal D-family DNA polymerase small subunit reveals evolutionary relationship to eukaryotic B-family polymerases. *FEBS Lett.* **584:** 3370-3375

Yao, M., Hatahet, Z., Melamede, R. J. and Kow, Y. W. (1994) Purification and characterization of a novel deoxyinosine-specific enzyme, deoxyinosine 3' endonuclease, from *Escherichia coli*. *J Biol Chem.* **269:** 16260-8

Yonekura, S., Nakamura, N., Yonei, S. and Zhang-Akiyama, Q. M. (2009) Generation, biological consequences and repair mechanisms of cytosine deamination in DNA. *J. Radiat. Res. (Tokyo).* **50:** 19-26

Youngblood, B., Bonnist, E., Dryden, D. T., Jones, A. C. and Reich, N. O. (2008) Differential stabilization of reaction intermediates: specificity checkpoints for M.EcoRI revealed by transient fluorescence and fluorescence lifetime studies. *Nucleic Acids Res.* **36:** 2917-25

Yuzhakov, A., Kelman, Z., Hurwitz, J. and O'Donnell, M. (1999) Multiple competition reactions for RPA order the assembly of the DNA polymerase delta holoenzyme. *EMBO J.* **18:** 6189-99

Zhang, Y., Yuan, F., Xin, H., Wu, X., Rajpal, D. K., Yang, D. and Wang, Z. (2000) Human DNA polymerase κ synthesizes DNA with extraordinarily low fidelity. *Nucleic Acids Res.* **28:** 4147–4156

Zharkov, D. O., Mechetin, G. V. and Nevinsky, G. A. (2010) Uracil-DNA glycosylase: structural, thermodynamic and kinetic aspects of lesion search and recognition. *Mutat. Res.*

Zheng, L. and Shen, B. (2011) Okazaki fragment maturation: nucleases take centre stage. *J. Mol. Cell Biol.* **3:** 23-30

Zhou, W., Rousset, F. and O'Neill, S. (1998) Phylogeny and PCR-based

classification of *Wolbachia* strains using *wsp* gene sequences. *Proc. Biol. Sci.* **265:** 509-515

Chapter 9

Appendices

Appendix A

Nucleic Acids Research Advance Access published September 26, 2009

Nucleic Acids Research, 2009, 1–9
doi:10.1093/nar/gkp800

The 3′–5′ proofreading exonuclease of archaeal family-B DNA polymerase hinders the copying of template strand deaminated bases

Henry J. Russell, Tomas T. Richardson, Kieran Emptage and Bernard A. Connolly*

Institute of Cell and Molecular Biosciences (ICaMB), University of Newcastle, Newcastle upon Tyne, NE2 4HH, UK

Received March 26, 2009; Revised August 25, 2009; Accepted September 11, 2009

ABSTRACT

Archaeal family B polymerases bind tightly to the deaminated bases uracil and hypoxanthine in single-stranded DNA, stalling replication on encountering these pro-mutagenic deoxynucleosides four steps ahead of the primer–template junction. When uracil is specifically bound, the polymerase–DNA complex exists in the editing rather than the polymerization conformation, despite the duplex region of the primer-template being perfectly base-paired. In this article, the interplay between the 3′–5′ proofreading exonuclease activity and binding of uracil/hypoxanthine is addressed, using the family-B DNA polymerase from *Pyrococcus furiosus*. When uracil/hypoxanthine is bound four bases ahead of the primer–template junction (+4 position), both the polymerase and the exonuclease are inhibited, profoundly for the polymerase activity. However, if the polymerase approaches closer to the deaminated bases, locating it at +3, +2, +1 or even 0 (paired with the extreme 3′ base in the primer), the exonuclease activity is strongly stimulated. In these situations, the exonuclease activity is actually stronger than that seen with mismatched primer-templates, even though the deaminated base-containing primer-templates are correctly base-paired. The resulting exonucleolytic degradation of the primer serves to move the uracil/hypoxanthine away from the primer–template junction, restoring the stalling position to +4. Thus the 3′–5′ proofreading exonuclease contributes to the inability of the polymerase to replicate beyond deaminated bases.

INTRODUCTION

DNA polymerases from many species, including bacteria, viruses and eukaryotes, possess a 3′–5′ proofreading exonuclease activity which removes misincorporated bases from extending primers, thereby improving fidelity. The catalytic centres responsible for polymerase and exonuclease functions are well separated and co-crystal structures with DNA being extended (polymerase conformation) or subject to exonucleolytic proofreading (editing conformation) are distinct (1–6). The family-B DNA polymerases from the archaeal domain also demonstrate proofreading activity, and, consequently, are able to synthesize DNA with high accuracy. This feature, combined with extreme thermostability, makes these enzymes very useful in the PCR (7,8). Crystal structures of a number of archaeal polymerases, e.g. from *Thermococcus gorgonarius* (Tgo-Pol) have demonstrated, as expected, significant distance between the polymerase and exonuclease active sites (9).

Archaeal DNA polymerases have an additional unusual property, binding tightly to the deaminated bases uracil and hypoxanthine and stalling DNA replication when these bases are encountered (10–16). During replication, these polymerases scan the template strand ahead of the replication fork and capture uracil/hypoxanthine in a specific pocket, when it is encountered four bases ahead of the primer–template junction (10,15). The subsequent cessation of replication stops the copying of uracil with adenine, and the conversion of a C:G to a T:A base-pair, in cases where the uracil resulted from cytosine deamination (10). Similarly, trapping of hypoxanthine (which can arise from adenine deamination) prevents A:T to G:C transitions. Recently, a crystal structure of Tgo-Pol, in complex with a primer-template containing uracil at the optimal +4 position in the template, has been solved (13). As anticipated (11), the uracil was flipped into a specific

*To whom correspondence should be addressed. Tel: +44 191 222 7371; Fax: +44 191 222 7424; Email: b.a.connolly@ncl.ac.uk

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

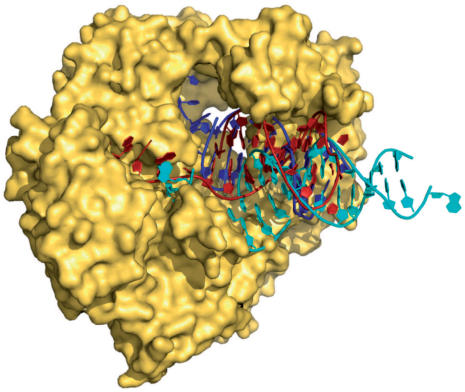


Figure 1. Structure of Tgo-Pol bound to a uracil-containing primer-template (red) (13). Superimposed are the expected positions of DNA bound in the polymerization mode (cyan) (17) and the editing mode (blue) (18) derived from structures of the family-B polymerase from bacteriophage RB69. The position of the uracil-containing primer-template clearly maps to the editing conformation more closely than to the polymerization.

binding pocket located within the polymerase N-terminal domain. Comparison with a related family B DNA polymerase from the RB-69 bacteriophage, for which both editing and polymerization complexes have been determined (17,18), indicated that the archaeal polymerase bound the uracil-containing DNA in an editing mode (Figure 1). This observation was somewhat unexpected as the duplex region of the uracil-containing primer-template contained only *bona fide* Watson–Crick base-pairs.

Prior to the structure of Tgo-Pol with a uracil-containing primer-template, no link between deaminated base recognition, mediated by the N-terminal domain, and proofreading activity, catalysed by a distinct exonuclease domain, was suspected. Indeed, most mechanistic and structural investigations of deaminated base recognition have been conducted with polymerase mutants lacking exonuclease activity (exo^-), to prevent unwanted degradation of primer-templates (10–15). With the observation that uracil capture places the DNA in an editing conformation, this study investigates any potential role of the 3′–5′ exonuclease activity in preventing replication beyond template strand deaminated bases. The family B polymerase from *Pyrococcus furiosus* (Pfu-Pol), which has ~80% amino acid sequence identity to Tgo-Pol and has previously been employed extensively to characterize uracil recognition, has been used in all experiments.

MATERIALS AND METHODS

Exonuclease assays

The primer-templates (primers labelled at the 5′-end with Cy5) used for exonuclease assays are shown in Table 1. Reactions were carried out in 400 µl of 20 mM Tris–HCl (pH 8.5), 10 mM KCl, 20 mM $MgSO_4$, 10 mM $(NH_4)_2SO_4$, 20 nM primer-template, 100 nM Pfu-PCNA (19) and 100 nM Pfu-Pol (20), the last component being added for initiation. Exonuclease assays were carried out with both wild type Pfu-Pol (20) and V93Q, a deaminated base insensitive mutant (11,13). The assay temperature was 30°C and timed 40 µl aliquots were withdrawn and the reaction quenched by addition of 40 µl stop buffer (40% formamide, 0.1 M EDTA and orange G) and 1 µl of a 100 µM solution of 'competitor DNA' (an exact complement of the template strand under study but lacking Cy5). The samples were denatured by heating to 90°C for 10 min and then rapidly cooled on ice. The excess of the 'competitor' prevents any rehybridization of the Cy5 primer to the template and ensures all the Cy5 primer, and products derived from it, remain single stranded during analysis. Products were detected using denaturing polyacrylamide (15%) gel electrophoresis (even while using denaturing gels, we observed significant hybridization of the Cy5 primer to the template if the competitor was omitted, the resulting double-stranded structures interfering with the assay) followed by fluorescence detection using Typhoon scanner (GE Healthcare). For reactions with fast time courses, an RQF-63 rapid quench flow apparatus was used (Hi-Tech Scientific, Bradford on Avon, UK).

Primer–template extension assays

The conditions and analysis methods were identical to exonuclease assays except that 400 µM each of dATP, dCTP, dGTP and dTTP were added to the reaction mixture and the assay temperature was 50°C. Extension assays were carried out both with wild type Pfu-Pol and the D215A point mutant, which is disabled in 3′–5′ exonuclease activity (20). A second set of experiments used the deaminated base insensitive mutation V93Q (11,13) and the double mutant V93Q/D215A, disabled in both deaminated base recognition and 3′–5′ exonuclease activity.

Data analysis

For band-density analysis of gel images, 'Image Quant' software (GE Healthcare) was used to determine the percentages of substrate and products. Data was fitted to the equation for a first order reaction using 'Grafit' (Erithacus Software, London), allowing determination of the rate constant for selected reactions.

RESULTS

Primer–templates for Pfu-Pol exonuclease assays

The primer-templates shown in Table 1 have been exploited to investigate any coupling between deaminated base recognition and 3′–5′ exonuclease by Pfu-Pol. The series labelled AA/TT has, for the control, two A:T base pairs in the double-stranded region immediately adjacent to the primer–template junction. Within the AA/TT set, the first two entries are primer-templates, containing only

Table 1. The primer-templates used in exonuclease and extension assays

Designation	Sequence
Control	Cy5-GGGGATCCTCTAGAGTCGACCTGCAGGGCAA
(AA/TT)	CCCCTAGGAGATCTCAGCTGGACGTCCCGTTCGTTCGAACAGAGG
Mismatch	Cy5-GGGGATCCTCTAGAGTCGACCTGCAGGGCAA
(AA/TT)	CCCCTAGGAGATCTCAGCTGGACGTCCCGT**C**CGTTCGAACAGAGG
U-1	Cy5-GGGGATCCTCTAGAGTCGACCTGCAGGGCAA
(AA/TT)	CCCCTAGGAGATCTCAGCTGGACGTCCCG**U**TCGTTCGAACAGAGG
U 0	Cy5-GGGGATCCTCTAGAGTCGACCTGCAGGGCAA
(AA/TT)	CCCCTAGGAGATCTCAGCTGGACGTCCCGT**U**CGTTCGAACAGAGG
U + 1	Cy5-GGGGATCCTCTAGAGTCGACCTGCAGGGCAA
(AA/TT)	CCCCTAGGAGATCTCAGCTGGACGTCCCGTT**U**GTTCGAACAGAGG
U + 2	Cy5-GGGGATCCTCTAGAGTCGACCTGCAGGGCAA
(AA/TT)	CCCCTAGGAGATCTCAGCTGGACGTCCCGTTC**U**TTCGAACAGAGG
U + 3	Cy5-GGGGATCCTCTAGAGTCGACCTGCAGGGCAA
(AA/TT)	CCCCTAGGAGATCTCAGCTGGACGTCCCGTTCG**U**TCGAACAGAGG
U + 4	Cy5-GGGGATCCTCTAGAGTCGACCTGCAGGGCAA
(AA/TT)	CCCCTAGGAGATCTCAGCTGGACGTCCCGTTCGT**U**CGAACAGAGG
Control	Cy5-GGGGATCCTCTAGAGTCGACCTGC
(GC/CG)	CCCCTAGGAGATCTCAGCTGGACGACCGTTCGTTCGAACAGAGG
Mismatch	Cy5-GGGGATCCTCTAGAGTCGACCTGC
(GC/CG)	CCCCTAGGAGATCTCAGCTGGAC**A**ACCGTTCGTTCGAACAGAGG
U + 2	Cy5-GGGGATCCTCTAGAGTCGACCTGC
(GC/CG)	CCCCTAGGAGATCTCAGCTGGACGA**U**CGTTCGTTCGAACAGAGG
U + 4	Cy5-GGGGATCCTCTAGAGTCGACCTGC
(GC/CG)	CCCCTAGGAGATCTCAGCTGGACGACC**U**TTCGTTCGAACAGAGG
H + 2	Cy5-GGGGATCCTCTAGAGTCGACCTGC
(GC/CG)	CCCCTAGGAGATCTCAGCTGGACGA**H**CGTTCGTTCGAACAGAGG
H + 4	Cy5-GGGGATCCTCTAGAGTCGACCTGC
(GC/CG)	CCCCTAGGAGATCTCAGCTGGACGACC**H**TTCGTTCGAACAGAGG

The top oligodeoxynucleotide serves as the primer strand (written in 5′–3′ direction) and the bottom as the template (3′–5′ direction). Emboldened and underlining is used to highlight bases that differ from the appropriate control template.

the four standard bases, either completely Watson–Crick base-paired (control), or with a single mismatch at the primer–template junction (mismatch). The remaining nucleic acids all contain a single uracil in the template strand. In two cases, the uracil is located in the duplex region [at positions −1 and 0 (at the primer–template junction)], in both instances being correctly base-paired with adenine. For the remainder, the uracil is located in the single-stranded region of the template spanning positions +1 to +4. A second series, documented GC/CG, has a slightly shorter primer strand which, with the control, places two G:C base pairs at the primer–template junction. This set consists of a fully base-paired control, a single mismatch at the primer–template junction and uracil or hypoxanthine at locations +2 and +4.

Exonuclease activity of Pfu-Pol in the presence of template strand deaminated bases

Any influence of the polymerase 3′–5′ exonuclease activity on the recognition of deaminated base was initially investigated using the AA/TT set under single turnover kinetic conditions. The polymerase binds tightly to uracil/hypoxanthine-containing primer-templates, resulting in complete binding of the DNA with a slight excess of protein at achievable concentrations of both components (15,19). Unfortunately, the poor interaction of the polymerase with standard primer-templates (e.g. DNA labelled control and mismatch in Table 1) makes full binding of the DNA difficult to achieve at reasonable concentrations of the two macromolecules. Fortunately, the presence of PCNA, the processivity clamp for Pfu-Pol, significantly improves the interaction of primer-templates with the polymerase. Therefore, all reactions contained Pfu-Pol (100 nM), PCNA (100 nM) and primer-template (20 nM), conditions previously shown to result in full binding of primer-templates, including those lacking uracil/hypoxanthine, to the polymerase and, hence, single turnover conditions (19). Figure 2 shows the exonucleolysis of the AA/TT primer-templates listed in Table 1, under the single turnover conditions detailed above, and over a time course of two minutes. The fully base-paired control is degraded with a half life of about 1 min and, as expected, the mismatched substrate is hydrolysed more rapidly, with most of the starting material removed at about 45 s. Remarkably, the primer-templates containing uracil at positions 0, +1, +2 and +3 are broken down very rapidly with the majority of the initial primer removed within 10 s. In contrast, when uracil is at +4, exonucleolysis is slow and most of the primer persists for the entire 2 min reaction time. The primer-template with uracil at −1 behaves in a similar manner to the control and is broken down neither especially rapidly nor slowly, as seen with the other uracil-containing substrates.

To confirm the generality of the above results, a second set of experiments has been undertaken with the GC/CG series of primer-templates. The essential difference is that the first set contains two A:T base-pairs at the

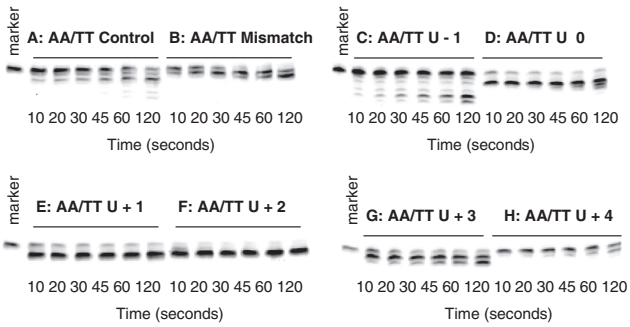


Figure 2. Exonuclease assay gel images for the AA/TT primer-templates listed in Table 1 observed with Pfu-Pol (exo$^+$). The primer-templates are identified above each of the panels. The marker is the primer itself, showing the position of the starting material.

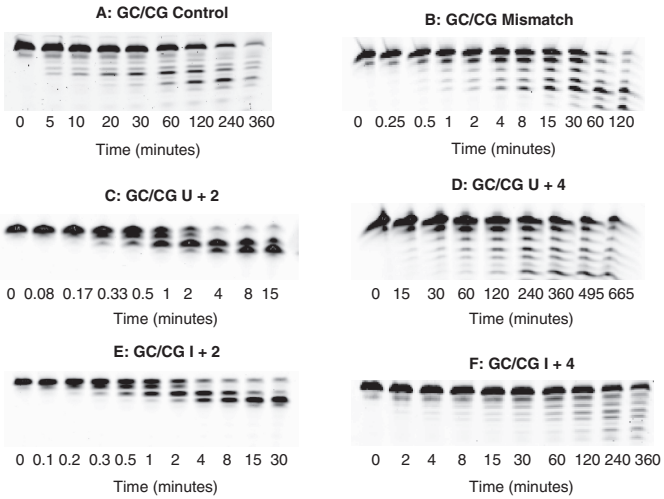


Figure 3. Exonuclease assay gel images for the GC/CG primer-templates listed in Table 1 observed with Pfu-Pol (exo$^+$). The primer-templates are identified above each panel. Time 0 is the observation prior to adding enzyme and gives the position of the starting materials.

primer–template junction, the second two G:Cs. The A:T series is expected to be more susceptible to 'fraying' than the G:C, giving rise, more easily, to single-stranded regions at the immediate 3′-terminus of the template. The proofreading 3′–5′ exonuclease activity of DNA polymerases requires unwinding of the template, with the generation of single strands, for activity (17,18). Therefore, the nature of the bases at the junction may influence the role the exonuclease plays on encountering deaminated bases. However, as shown in Figure 3, the exonuclease acts similarly on uracil, regardless of the base composition at the primer–template junction. Thus, for the GC/CG set the presence of uracil at +2 results in rapid exonucleolyis, faster than that observed for the mismatch. With uracil at +4, hydrolysis is slightly slowed, relative to the fully base-paired control. Uracil at other positions (U+1, U+3) causes the polymerase to behave in an analogous manner to that seen with the AA/TT primer-templates, with rapid exonuclease activity (data not shown). Figure 3 also shows the behaviour seen with hypoxanthine, the deamination product of adenine, which is also recognized by the polymerase (15,19). As with uracil, the presence of hypoxanthine at the +2 and +4 locations results in very rapid and marginally reduced exonucleolytic degradation, respectively.

To put the data observed in Figures 2 and 3 on a more quantitative basis, the first order rate constants for the exonucleolysis of selected primer-templates have been

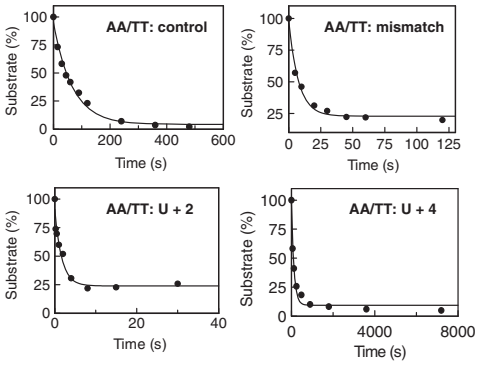


Figure 4. Determination of the rate constants for the exonucleolysis of primer-templates. Data, obtained from gels similar to those shown in Figures 2 and 3, were fitted to a first order decay to show the disappearance of substrate. Only the data obtained with the AA/TT series is shown (similar quality fits were obtained with the GC/CG set). For U+2, a rapid quench apparatus was used; for the other three primer-templates, manual stopping of the reactions was sufficient. The rate constants found are given in Table 2.

determined. For the AA/TT set, experiments were carried out with the control, mismatch, U+2 (as representative of U0, +1, +2 and +3) and U+4 primer-templates. The same representatives (control, mismatch U+2 and U+4) were used with the GC/CG primer-templates, and in this case, H+2 and H+4 were also investigated. Rate constants were determined by carrying out the hydrolysis reactions over appropriate time spans, and the gels obtained (not shown) were of similar quality to those in Figures 2 and 3. Fits to single exponential decay plots, obtained for the AA/TT series, are shown in Figure 4 and the first order rate constants obtained are summarized in Table 2. All members of the AA/TT series are degraded more rapidly than their corresponding GC/CG partners, presumably because the 'weaker' primer–template junction is more prone to 'fraying' and the formation of single strands needed for exonucleolysis. Thus, for example, the control AA/TT has a rate constant for degradation some 75-fold higher than the corresponding GC/CG control. Similar ratios are obtained from other comparable pairs and, therefore, the influence of both mismatches and deaminated bases are similar in the two sets. The mismatched primer-template is hydrolysed about 5 (GC/CG)–10 (AA/TT) fold faster than the control. Most striking is the very rapid degradation of primer-templates containing a deaminated base at +2, maintained for both GC/CG and AA/TT and occurring with both uracil and hypoxanthine. Depending on the precise primer-template, exonucleolysis takes place about two orders of magnitude faster than controls, at even more rapid rates than seen for mismatches. Sequences containing a deaminated base at +4 are hydrolysed more slowly than the appropriate control, albeit by rather small (<2) factors.

Table 2. The rate constants, determined under single turnover conditions, for the 3′–5′ proofreading exonuclease of Pfu-Pol with different primer-templates

Primer-template[a]	Rate constant[b] (min^{-1})	Rate (relative to appropriate control)
Control (AA/TT)	0.84 ± 0.06	1
Mismatch (AA/TT)	9 ± 2	10.7
U+2 (AA/TT)	48 ± 12	57
U+4 (AA/TT	0.42 ± 0.06	0.5
Control (GC/CG)	0.011 ± 0.001	1
Mismatch (GC/CG)	0.061 ± 0.007	5.5
U+2 (GC/CG)	1.8 ± 0.3	163
U+4 (GC/CG)	0.008 ± 0.0007	0.7
H+2 (GC/CG)	1.5 ± 0.3	136
H+4 (GC/CG)	0.009 ± 0.004	0.8

[a]The full primer–template sequences are given in Table 1.
[b]Each rate constant is the average of three values ± SD.

Primer–template extension by Pfu-Pol exo$^+$ and exo$^-$ with template strand deaminated bases

To determine the influence of the 3′–5′ exonuclease activity on the ability of Pfu-Pol to replicate beyond template strand deaminated bases, extension assays have been carried out with the control and U+4 primer-templates, in both the AA/TT and GC/CG series, listed in Table 1. Experiments were also recorded with I+4 in the GC/CG context. In contrast to the exonuclease experiments (performed at 30°C), extensions were carried out at 50°C in order to produce measurable results with deaminated base-containing templates. The results are given in Figure 5, which shows that replication using control templates (i.e. lacking deaminated bases) is the same for the exo$^+$ and exo$^-$ variants of Pfu-Pol (Figure 5, panels A and C). The extension ladders are identical with full length product appearing in 5 min. In agreement with earlier results (10,15,19), copying beyond uracil is extremely slow and requires extended times of up to 24 h. Nevertheless, it is apparent that the exo$^-$ variant is more proficient at extension, with product clearly visible at 1 and 3 h and most of the starting material consumed after 6 h. With the wild type exo$^+$ Pfu-Pol barely any product can be seen after 1 and 3 h and the majority of the substrate is still present after 6 h (Figure 5, panels B and D). Similar results were seen with I+4 (Figure 5, panel E), where the exo$^-$ variant produced more full length product then exo$^+$.

Stimulation of exonuclease activity is dependent on deaminated base binding

To verify that the marked increase in exonuclease rates seen on encountering uracil/hypoxanthine is a consequence of their specific binding, a deaminated base insensitive mutant, V93Q, has been used. Valine 93 makes a hydrophobic stacking interaction with uracil and its replacement with glutamine sterically occludes binding (13). As a result, V93Q does not bind strongly to deaminated bases and the stalling response is largely abolished. As shown in Figure 6 (panels labelled with an A),

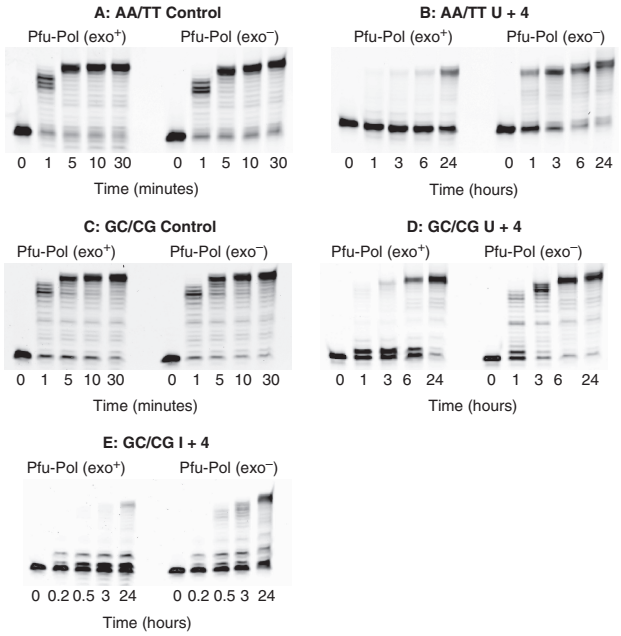


Figure 5. Gel images for the extension of primer-templates using wild type Pfu-Pol (exo$^+$) and Pfu-Pol D215A, a 3′–5′ proofreading exonuclease deficient mutant (exo$^-$). Unlike exonuclease assays (performed at 30°C), extensions were carried out at 50°C in order to produce observable incorporation with deaminated base-containing templates. The primer-templates are identified above each image. Note the much longer time courses needed to observe extension with deaminated bases present as compared to controls.

the gel patterns seen for exonucleolysis of the GC/CG set are largely independent of the presence of a deaminated bases. Confirmation comes from Table 3, which illustrates that there is little difference in the rate constants observed for the exonuclease activity of V93Q with the control, U + 2, U + 4, H + 2 and H + 4. In particular, there is no marked acceleration with a deaminated base at + 2, as seen with the wild type polymerase. A slight (2.6-fold) increase is seen with V93Q and U + 2, but this is very much less than the factors measured with the wild type and may be accounted for by residual binding ability of the mutant. It is also noted that V93Q has a less powerful endonuclease activity than the wild type (compare the GC/CG controls in Tables 2 and 3); at present, the origin of this difference is unclear. Figure 6 (panels marked B) also shows that the exo$^-$ variant of V93Q is no more proficient at reading beyond deaminated bases than exo$^+$, again in contrast to the wild type. V93Q exo$^-$ is slightly better at extension than exo$^+$ (again why this arises is unclear), but this is a general change seen with the controls as well as the uracil/hypoxanthine containing primer-templates.

CONCLUSIONS

This article elucidates the role that the 3′–5′ proofreading exonuclease activity of archaeal family-B DNA polymerases plays in uracil recognition. Control experiments established the background rates for fully base-paired primer-templates and mismatch substrates. The increase in rate constant by about an order of magnitude when a mismatch is present at the primer–template junction is typical for DNA polymerases (3,21,22). However, we note that primer-templates terminated in two A:T base pairs are acted on much more rapidly than those ending G:C, assuredly due to the greater ease of unwinding the former. Using a fully base-paired primer-template with uracil or hypoxanthine at the + 4 position results in slight inhibition of the exonuclease activity and such nucleic acids are also extended extremely slowly (10–12,14,15,19). A crystal structure of Tgo-Pol, complexed with a U + 4 primer-template, shows that the DNA sequence adopts an editing conformation but the 3′ base of the primer remains double-stranded and does not protrude into the exonuclease catalytic centre (Figure 1) (13). A primer-template bound in such a mode would be acted upon poorly by both the exonuclease and polymerase activities, explaining the observed results. Perhaps the most remarkable finding is the strong stimulation of exonuclease activity, by about two orders of magnitude, with primer-templates containing uracil at + 3, + 2, + 1 and 0 and also with hypoxanthine at + 2. No structural data are available for an archaeal

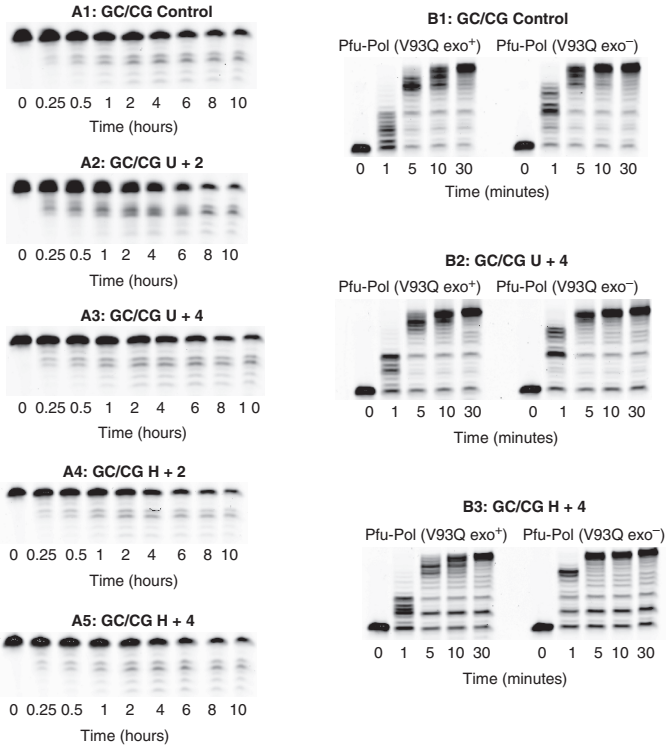


Figure 6. Gel images for the exonucleolyis and polymerization by a deaminated base insensitive mutation, Pfu-Pol V93Q. The figures on the left (designated with an A) are 3′–5′ exonuclease assays seen with V93Q for the primer-templates identified on the top of each of the panels. The rate constants obtained from these gels are given in Table 3. The figures on the right (designated with a B) are polymerization assays for both V93Q exo$^+$ and V93Q exo$^-$ using the primer-templates identified above each of the panels. All assays were carried out at 30°C.

Table 3. The rate constants, determined under single turnover conditions, for the 3′–5′ proofreading exonuclease of Pfu-Pol V93Q with different primer-templates

Primer-template[a]	Rate constant[b] (min^{-1})	Rate (relative to control)
Control (GC/CG)	0.0026 ± 0.0006	1
U + 2 (GC/CG)	0.0067 ± 0.002	2.6
U + 4 (GC/CG)	0.0034 ± 0.001	1.3
H + 2 (GC/CG)	0.003 ± 0.0003	1.1
H + 4 (GC/CG)	0.0027 ± 0.0001	1

[a]The full primer–template sequences are given in Table 1.
[b]Each rate constant is the average of three values ± SD.

polymerase in complex with DNA containing a deaminated base at any of these locations. We would suggest that the DNA binds in an editing conformation but, additionally, the 3′-terminus of the primer strand unwinds to give a single-stranded region that enters the exonuclease active site (Figure 7). Although archaeal polymerases bind most tightly to primer-templates containing uracil at +4, significant affinity is still seen when uracil is positioned at +3, +2 and +1 (12). Previously it was unclear how the enzyme could accommodate shifts in uracil location without significant relative re-positioning of the uracil-binding pocket and the active site amino acids responsible for interaction with the primer–template junction. Linking the degree of melting of the primer strand with the position of uracil would maintain the effective separation between uracil and the primer–template junction at around 4 bases. It seems clear that the binding energy available from the polymerase–uracil/hypoxanthine interaction can, in certain circumstances, result in the unwinding of fully base-paired primer-templates, leading to rapid exonuclease activity. We note that the stimulation of the exonuclease ceases with uracil at −1 and the polymerase is no longer able to recognize the base once this position is reached. Further, the effect is observed for both 'weak' and 'strong' primer-templates, terminated in A:T and G:C base-pairs, respectively.

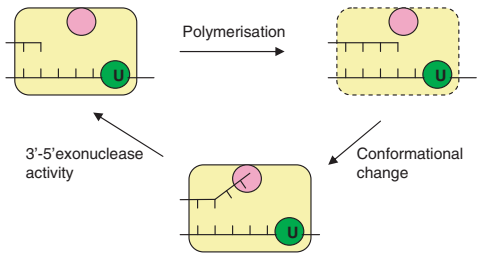


Figure 7. Idling by Pfu-Pol on encountering a template strand deaminated base. The polymerase (yellow) captures uracil four bases ahead of the primer–template junction using a specific binding pocket (green). Further, extremely slow, polymerization can add two bases to give a strained conformation (indicated by hatched borders) in which the primer–template junction and the uracil binding site are too close, with a separation of two bases. A conformational change restores the spacing to four bases by unwinding the terminal two bases in the primer and, hence, relieves the strain. This conformational change places the single-stranded bases in the 3′–5′ exonuclease site (rose) enabling their removal.

Thus interaction of the polymerase with deaminated bases appears to be powerful enough to unwind even robustly base-paired junction.

It is well established that the main role of the 3′–5′ proofreading exonuclease activity is to excise bases aberrantly incorporated by the polymerase, thereby correcting replication errors and increasing fidelity (1–6). Additionally, the exonuclease appears to play a role in handling any non-canonical bases that a polymerase may encounter on template strands during replication. Many polymerases have been observed to 'idle' on running into template strand damage, making sequential use of their polymerase and exonuclease activities to repeatedly add and remove a standard base opposite the lesion (3,22–25). During 'idling', repeated dNTP to dNMP turnover is not accompanied by net DNA synthesis or degradation. Critically, the polymerase is prevented from progressing beyond the damage and the potential for permanently fixing a mutation is, thereby, avoided. It has also been suggested that turnover of dNTPs may be the signal for initiation of appropriate repair pathways (23). Previous studies have suggested that deaminated base recognition, a function unique to family-B archaeal replicative DNA polymerases (26), acts to proofread the template strand and prevent replication beyond uracil/hypoxanthine (10–14). Uracil can arise in DNA by deamination of cytosine in C:G base-pairs to give a pro-mutagenic U:G mispair, a process favoured by the hyperthermophilic environments of many archaea (27–29). Uracil is mistaken for thymine by most DNA polymerases and unless the U:G mismatch is repaired, replication results in 50% of the progeny inheriting a C:G → T:A transition mutation. Similarly, deamination of adenine to hypoxanthine results in A:T → G:C transitions following replication. Stalling of replication by the archaeal polymerase prevents such deaminated base-induced mutations and presumably also initiates DNA repair by an, as yet unknown pathway, that probably involves error-free recombination (30,31). This article shows that the 3′–5′ exonuclease activity of archaeal polymerases plays a role in preventing replication beyond uracil/hypoxanthine, further checking the template strand proofreading function of the deaminated base-binding pocket. During replication, stalling takes place when uracil/hypoxanthine is encountered at +4, largely switching off both the polymerase and exonuclease activities. Any further progression of the polymerase towards the deaminated base results in strong activation of the exonuclease activity, which degrades and shortens the extending primer (Figure 7). As a consequence, the separation between the deaminated base and the primer–template junction is restored, strongly decreasing the probability that the polymerase proceeds beyond uracil/hypoxanthine. The influence of the exonuclease activity is apparent from Figure 5; an exo$^-$ variant gives more full-length product in the presence of uracil than the wild type enzyme with a functional proofreading activity. Repeated polymerase/exonuclease cycles in response to deaminated bases is reminiscent of 'idling' described above. However, until now, triggering of exonuclease activity, essential for 'idling', has been dependent on the mismatched base-pair produced when a polymerase copies a damaged base. Such mismatches both stimulate exonuclease activity and are difficult to extend (3,23). The archaeal polymerase response to uracil/hypoxanthine does involve a damaged base, but not a base-pair mismatch as the deaminated base remains in the single-stranded region of the template. Rather, 'idling' in this instance flows from the specific capture of uracil/hypoxanthine by the N-terminal domain of the polymerase (Figure 7).

ACKNOWLEDGEMENTS

The authors thank Pauline Heslop for providing expert technical assistance.

FUNDING

UK Biotechnology and Biological Sciences Research Council [grant no. BBS/B/0560 to B.A.C]. Funding for open access charge: UK Biotechnology and Biological Research Council (BBSRC).

Conflict of interest statement. None declared.

REFERENCES

1. Joyce,C.M. (1989) How DNA travels between the separate polymerase and 3′–5′ exonuclease sites of DNA polymerase I (Klenow fragment). *J. Biol. Chem.*, **264**, 10858–10866.
2. Kunkel,T.A. (1992) DNA replication fidelity. *J. Biol. Chem.*, **267**, 18251–18254.
3. Johnson,K.A. (1993) Conformational coupling in DNA polymerase fidelity. *Annu. Rev. Biochem.*, **62**, 685–713.
4. Joyce,C.M. and Steitz,T.A. (1994) Function and structure relationships in DNA polymerases. *Annu. Rev. Biochem.*, **63**, 777–822.
5. Benkovic,S.J., Valentine,A.M. and Salinas,S. (2001) Replisome-mediated DNA replication. *Annu. Rev. Biochem.*, **70**, 181–208.

6. Joyce,C.M. and Benkovic,S.J. (2004) DNA polymerase fidelity: kinetics, structure, and checkpoints. *Biochemistry*, **43**, 14317–14324.
7. Lundberg,K.S., Shoemaker,D.D., Adams,M.W.W., Short,J.M., Sorge,J.A. and Mathur,E.J. (1991) High-fidelity amplification using a thermostable DNA polymerase isolated from *Pyrococcus furiosus*. *Gene*, **101**, 1–6.
8. Cline,J., Braman,J.C. and Hogrefe,H.H. (1996) PCR fidelity of *Pfu* DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Res.*, **24**, 3546–3551.
9. Hopfner,K.-P., Eichinger,A., Engh,R.A., Laue,F., Ankenbauer,W., Huber,R. and Angerer,B. (1999) Crystal structure of a thermostable type B DNA polymerase from *Thermococcus gorgonarius*. *Proc. Natl Acad. Sci. USA*, **96**, 3600–3605.
10. Greagg,M.A., Fogg,M.J., Panayotou,G., Evans,S.J., Connolly,B.A. and Pearl,L.H. (1999) A read-ahead function in archaeal DNA polymerases detects promutagenic template-strand uracil. *Proc. Natl Acad. Sci. USA*, **96**, 9045–9050.
11. Fogg,M.J., Pearl,L.H. and Connolly,B.A. (2002) Structural basis for uracil recognition by archaeal family B DNA polymerases. *Nat. Struct. Biol.*, **9**, 922–927.
12. Shuttleworth,G., Fogg,M.J., Kurpiewski,M.R., Jen-Jacobson,L. and Connolly,B.A. (2004) Recognition of the pro-mutagenic base uracil by family B DNA polymerases from archaea. *J. Mol. Biol.*, **337**, 621–634.
13. Firbank,S.J., Wardle,J., Heslop,P., Lewis,R.J. and Connolly,B.A. (2008) Uracil recognition in archaeal DNA polymerases captured by X-ray crystallography. *J. Mol. Biol.*, **381**, 529–539.
14. Connolly,B.A. (2009) Recognition of deaminated bases by archaeal family-B polymerases. *Biochem. Soc. Trans.*, **37**, 65–68.
15. Gill,S., O'Neill,R., Lewis,R.J. and Connolly,B.A. (2007) Interaction of the family-B DNA polymerase from the archaeon *Pyrococcus furiosus* with deaminated bases. *J. Mol. Biol.*, **372**, 855–863.
16. Gruz,P., Shimizu,M., Pisani,F.M., DeFelice,M., Kanke,Y. and Nohmi,T. (2003) Processing of DNA lesions by archaeal DNA polymerases from *Sulfolobus solfataricus*. *Nucleic Acids Res.*, **31**, 4024–4030.
17. Franklin,M.C., Wang,J. and Steitz,T.A. (2001) Structure of the replicating complex of a Pol alpha family DNA polymerase. *Cell*, **105**, 657–666.
18. Shamoo,Y. and Steitz,T.A. (1999) Building a replisome from interacting pieces: sliding clamp complexed to a peptide from DNA polymerase and a polymerase editing complex. *Cell*, **99**, 155–166.
19. Emptage,K., O'Neill,R., Solovyova,A. and Connolly,B.A. (2008) Interplay between DNA polymerase and proliferating cell nuclear antigen switches off base excision repair of uracil and hypoxanthine during replication in archaea. *J. Mol. Biol.*, **383**, 762–771.
20. Evans,S.J., Fogg,M.J., Mamone,A., Davis,M., Pearl,L.H. and Connolly,B.A. (2000) Improving dideoxynucleotide-triphosphate utilisation by the hyper-thermophilic DNA polymerase from *Pyrococcus furiosus*. *Nucleic Acids Res.*, **28**, 1059–1066.
21. Donlin,M.J., Patel,S.S. and Johnson,K.A. (1991) Kinetic partitioning between the exonuclease and polymerase sites in DNA error correction. *Biochemistry*, **30**, 538–546.
22. Khare,V. and Eckert,K.A. (2001) The 3′–5′ exonuclease of T4 DNA polymerase removes premutagenic alkyl mispairs and contributes to futile cycling at 0^6 methylguanine lesions. *J. Biol. Chem.*, **276**, 24286–24292.
23. Khare,V. and Eckert,K.A. (2002) The proofreading 3′–5′ exonuclease activity of DNA polymerases: a kinetic barrier to translesion DNA synthesis. *Mutation Res.*, **510**, 45–54.
24. Schwartz,H., Shavitt,O. and Livneh,Z. (1988) The role of exonucleolytic processing and polymerase-DNA association in vitro: significance for SOS targeted mutagenesis. *J. Biol. Chem.*, **263**, 18277–18285.
25. Garg,P., Stith,C.M., Sabouri,N., Johansson,E. and Burgers,P.M.J. (2004) Idling by DNA polymerase delta maintains a ligatable nick during lagging-strand DNA replication. *Genes Dev.*, **18**, 2764–2773.
26. Wardle,J., Burgers,P.M.J., Cann,I.K.O., Darley,K., Heslop,P., Johansson,E., Lin,L., McGlynn,P., Sanvoisin,J., Stith,C.M. *et al.* (2008) Uracil recognition by replicative DNA polymerases is limited to the archaea, not occurring with bacteria and eukarya. *Nucleic Acids Res.*, **36**, 705–711.
27. Lindahl,T. (1993) Instability and decay of the primary structure of DNA. *Nature*, **286**, 709–715.
28. Lindahl,T. and Nyberg,B. (1974) Heat-induced deamination of cytosine residues in deoxyribonucleic acid. *Biochemistry*, **13**, 3405–3410.
29. Schroeder,G.K. and Wolfenden,R. (2007) Rates of spontaneous disintegration of DNA and the rate enhancements produced by DNA glycosylases and deaminases. *Biochemistry*, **46**, 13638–13647.
30. Baynton,K. and Fuchs,R.P.P. (2000) Lesions in DNA: hurdles for polymerases. *Trends Biochem. Sci.*, **25**, 74–79.
31. Cox,M.M. (2001) Recombinational DNA repair of damaged replication forks in *Escherichia coli*: questions. *Annu. Rev. Genet.*, **35**, 53–82.

Appendix B: Full TK0664 sequence analysis from 2010

Sample #	TR1	TR2	TR3	TR4
Ai	-	g>a transition 9 bases upstream from start codon		g deletion at 200
Aii	g deletion at 58	g>a transition 9 bases upstream from start codon		g>a transition at 435 (= W145STOP)
Aiii	-	g>a transition 9 bases upstream from start codon	a deletion at 500	g>t transversion 10 bases upstream from start codon
Aiv	a insertion at 512	g>a transition 9 bases upstream from start codon	c deletion at 391	incomplete sequence...no mutation found
Av	t>c transition at 433 (= W145R)	g>a transition 9 bases upstream from start codon	g deletion at 125	g>t transversion 10 bases upstream from start codon
Avi	t>c transition at 433 (= W145R)	g insertion at 117	c insertion at 59	g>a transition at 263 (= G88E)
Avii	a>g transition at 464 (= D155G)	g>a transition 9 bases upstream from start codon	a deletion at 500	g>t transversion 10 bases upstream from start codon
Aviii	-	g>a transition 9 bases upstream from start codon	c deletion at 391	a deletion at 269
xxv	a>g transition at 464 (= D155G)		a deletion at 500	g deletion at 200
xxvi	a>g transition at 464 (= D155G)		g>a transition at 50 (= W17STOP)	g>a transition at 435 (= W145STOP)
i	t>g transversion at 433 (= W145G)	g>a transition 9 bases upstream from start codon	g>a transition at 283 (= D95N)	a deletion at 269
ii	t>g transversion at 433 (= W145G)	g>a transition 9 bases upstream from start codon	g insertion at 450	incomplete sequence...no mutation found
iii	10 base deletion	g>a transition 9 bases upstream from start codon	g>a transition at 283 (= D95N)	g>a transition at 435 = W145STOP
iv	a insertion at 573	g deletion at 190	g>a transition at 283 (= D95N)	a deletion at 269
v	g deletion at 124	incomplete sequence...no mutation found	6 base deletion containing start codon	a deletion at 269
vi	g deletion at 124	g>a transition 9 bases upstream from start codon	g>a transition at 283 (= D95N)	a deletion at 269
vii	incomplete sequence...no mutation found	g>a transition 9 bases upstream from start codon	g>a transition at 283 (= D95N)	a deletion at 269
viii	a deletion at 269	g>a transition 9 bases upstream from start codon	g>a transition at 283 (= D95N)	incomplete sequence...no mutation found
xxvii	a deletion at 473		g>a transition at 283 (= D95N)	a deletion at 269
xxviii	g deletion at 124		g>a transition at 283 (= D95N)	incomplete sequence...no mutation found
ix	a deletion at 498	g>a transition 9 bases upstream from start codon	t>c transition at 433 (= W145R)	incomplete sequence...no mutation found
x	a deletion at 326	g>a transition 9 bases upstream from start codon	g>a transition at 51 (= W17STOP)	g>a transition at 435 (= W145STOP)
xi	48 base insertion at 165	incomplete sequence...no mutation found	g>a transition at 51 (= W17STOP)	a deletion at 269
xii	31 base insertion at 217	g>a transition 9 bases upstream from start codon	g>a transition at 51 (= W17STOP)	g>a transition at 428 = (W143STOP)
xiii	49 base insertion at 165	g>a transition 9 bases upstream from start codon	g>a transition at 51 (= W17STOP)	incomplete sequence...no mutation found
xiv	incomplete sequence...no mutation found	g>a transition 9 bases upstream from start codon	incomplete sequence...no mutation found	incomplete sequence...no mutation found
xv	g deletion at 430	g>a transition 9 bases upstream from start codon	g>a transition at 51 (= W17STOP)	g>a transition at 69 (= W23STOP)
xvi	g>a transition at 90 (= W30STOP)	g>a transition 9 bases upstream from start codon	g>a transition at 51 (= W17STOP)	g>a transition at 435 (= W145STOP)
xxix	g>a transition at 90 (= W30STOP)		g>a transition at 51 (= W17STOP)	g deletion at 283
xxx	a deletion at 269		g>a transition at 51 (= W17STOP)	g>a transition at 435 (= W145STOP)

xvii	a deletion at 224	g>a transition 9 bases upstream from start codon	a deletion at 269	a deletion at 269
xviii	g>a transition at 429 (= W143STOP)	g>a transition 9 bases upstream from start codon	a deletion at 269	a deletion at 269
xix	g deletion at 428	g>a transition 9 bases upstream from start codon	incomplete sequence...no mutation found	a deletion at 269
xx	g deletion at 450	g>a transition 9 bases upstream from start codon	incomplete sequence...no mutation found	a deletion at 269
xxi	g deletion at 187	incomplete sequence...no mutation found	g insertion at 560	g>a transition at 435 = (W145STOP)
xxii	g deletion at 187	g>a transition 9 bases upstream from start codon	incomplete sequence...no mutation found	a>t transversion at 382 (= R128STOP)
xxiii	g deletion at 450	g>a transition 9 bases upstream from start codon	g deletion at 125	g>a transition at 435 (= W145STOP)
xxiv	cg insertion at 66	incomplete sequence...no mutation found	g deletion at 125	g>a transition at 435 (= W145STOP)
xxxi	g deletion at 187		g insertion at 560	g deletion at 201
xxxii	g insertion at 510		a deletion at 269	a deletion at 269
xxxiii	a deletion at 269		incomplete sequence...no mutation found	a deletion at 47
xxxiv	a deletion at 269		10 base deletion at 54	a deletion at 47
xxxv	g deletion at 450		g deletion at 125	a deletion at 47
xxxvi	g deletion at 583		g deletion at 120	g deletion at 201
xxxvii	g insertion at 265		incomplete sequence...no mutation found	a deletion at 47
xxxviii	incomplete sequence...no mutation found		wild type	a deletion at 47
xxxix	g>a transition at 429		g deletion at 120	incomplete sequence...no mutation found
xl	g insertion at 584		g>a transition at 197 (= W66STOP)	a deletion at 47
xli	a insertion at 175		g deletion at 200	g>a transition at 435 (= W145STOP)
xlii	a insertion at 175		g deletion at 200	g>a transition at 435 (= W145STOP)
xliii	c insertion at 519		g deletion at 161	g>a transition at 435 (= W145STOP)
xliv	a deletion at position 269		g deletion at 200	g>a transition at 435 (= W145STOP)
xlv	a deletion at 269		incomplete sequence...no mutation found	g>a transition at 435 (= W145STOP)
xlvi	cag deletion at 287		g deletion at 200	a deletion at 269
xlvii	c deletion at 395		g deletion at 200	g>a transition at 435 (= W145STOP)
xlviii	a deletion at 269		incomplete sequence...no mutation found	a deletion at 269
xlix	wild type		g deletion at 125	incomplete sequence...no mutation found
l	a deletion at 269		g>t transversion at 283 (= D95Y)	g deletion at 125
li	incomplete sequence...no mutation found		g deletion at 125	g deletion at 125
lii	c deletion at 395		incomplete sequence...no mutation found	incomplete sequence...no mutation found
liii			incomplete sequence...no mutation found	a insertion at 645
liv			g>t transversion at 283 (= D95Y)	incomplete sequence...no mutation found
lv			c deletion at 433	incomplete sequence...no mutation found

Ivi			g>t transversion at 283 (= D95Y)	incomplete sequence...no mutation found

Key: Black = unique mutation, green = sister/daughter mutation, yellow = incomplete sequence, Red = anomaly

Appendix C: Full TK0664 sequence analysis from 2011

Sample #	TR1	TR2	TR3	TR4
Ai	a insertion at 72	g deletion at 271	t>g transversion at 433 (= W145G)	inconclusive
Aii	g deletion at 430	a deletion at 269	g>t transversion at 190 (= E64STOP)	g>t transversion at 412 (= G138STOP)
Aiii	a insertion at 72	polyclonal DNA sequence/contamination	g>a transition at 295 (= D99N)	g>t transversion at 412 (= G138STOP)
Aiv	g deletion at 122	g insertion at 209	t>a transversion at 290 (= I97N)	g>a transition at 129 (= G43E)
Av	a insertion at 453	g>a transition at 429 (= W143STOP)		wild type
Avi	g deletion at 430	a deletion at 269		g>a transition at 50 (= W17STOP)
Avii	a insertion at 72	g>a transition at 198 (= W66STOP)		
Aviii	g deletion at 430	a deletion at 269		
Bi	g>a transition at 460 (= E>K)	a deletion at 269	t>g transversion at 433 (= W145G)	g insertion at 133
Bii	g>a transition at 460 (= E>K)	t>c transition at 433 (= W145R)	t>g transversion at 433 (= W145G)	a deletion at 326
Biii		a>t transversion at 600 (= E200D)	c deletion at 392	a deletion at 326
Biv	g>a transition at 460 (= E>K)	g deletion at 200	t>g transversion at 433 (= W145G)	a deletion at 326
Bv	g>a transition at 460 (= E>K)	c deletion at 655		a deletion at 326
Bvi	incomplete	g>a transition at 197 (= W66STOP)	g>a transition at 295 (= D99N)	a deletion at 326
Bvii	g>a transition at 460 (= E>K)	g>t transversion at 122 (= G41V)	t>g transversion at 433 (= W145G)	
Bviii	g>a transition at 460 (= E>K)	a deletion at 269	t>g transversion at 433 (= W145G)	
Bix	t insertion at 144	a deletion at 269		
Ci	a deletion at 269	t>c transition at 433 (= W145R)	g>a transition in start codon	g>a transition at 198 (= W66STOP)
Cii	a deletion at 269	polyclonal DNA sequence/contamination	g>a transition at 198 (= W66STOP)	g>a transition at 198 (= W66STOP)
Ciii	a deletion at 269	g>t transversion at 124 = G42C	g>a transition at 198 (= W66STOP)	g>a transition at 198 (= W66STOP)

Civ	17 base insertion at 176	a deletion at 168	g>a transition at 198 (= W66STOP)	g>a transition at 198 (= W66STOP)
Cv	16 base deletion at 557	t>c transition at 433 (= W145R)	g>a transition at 198 (= W66STOP)	g>a transition at 198 (= W66STOP)
Cvi	wild type	g insertion at 131	g>a transition at 198 (= W66STOP)	
Cvii	a deletion at 269	g insertion at 161		
Cviii	a deletion at 269			
Di	t>c transition at 433 (= W145R)	a deletion at 269	g deletion at 583	t>a transversion 433 (= W145R)
Dii	t>c transition at 433 (= W145R)	t>c transition at 433 (= W145R)	g>a transition at 198 (= W66STOP)	
Diii	t>c transition at 433 (= W145R)	a deletion at 269	g>a transition at 198 (= W66STOP)	
Div	t>c transition at 433 (= W145R)		g>a transition at 434 (= W145STOP)	
Dv	g deletion at 502		g>a transition at 198 (= W66STOP)	
Dvi	t>c transition at 433 (= W145R)			
Ei		a deletion at 269	at deletion at 98	t>g transversion at 433 (= W145G)
Eii		g>a transition at 449 (= W150STOP)		
Eiii		t>c transition at 433 (= W145R)		
Eiv		g deletion at 69		
Ev		g>a transition at 449 (= W150STOP)		
Evi		g>a transition at 449 (= W150STOP)		
Fi	g deletion at 59	g>a transition at 32 (= W11STOP)	g>t transversion 10 bases upstream from start codon	polyclonal
Fii	t>c transition at 433 (= W145R)	a deletion at 269		
Fiii	t>c transition at 433 (= W145R)	g deletion at 200		
Fiv	t>c transition at 433 (= W145R)	g deletion at 200		
Fv	t>c transition at 433 (= W145R)	a deletion at 269		
Fvi	t>c transition at 433 (= W145R)			
Fvii	t>c transition at 433 (= W145R)			
Gi	g>a transition at 198 (= W66STOP)	t>c transition at 148 (= C50R)	g>a transition at 198 (= W66STOP)	c>a transversion at 252 (= Y84STOP)
Gii	g>a transition 9 nucleotides upstream of start codon	g>a transition at 198 (= W66STOP)		
Giii	g>a transition 9 nucleotides upstream of start codon	Inconclusive		
Giv	g>a transition 9 nucleotides upstream of start codon	Inconclusive		
Gv	g>a transition 9 nucleotides upstream of start codon	Inconclusive		
Gvi	t>c transition at 433 (= W145R)	Inconclusive		
Gvii	g>a transition at 198 (= W66STOP)			
Gviii	g>a transition 9 nucleotides upstream of start codon			

Hi	t>c transition at 148 (= C50R)	t>c transition at 433 (= W145R)	incomplete…resequence	a deletion at 269
Hii	t>c transition at 148 (= C50R)	a deletion at 269		a deletion at 269
Hiii	t>c transition at 148 (= C50R)	Inconclusive		a deletion at 269
Hiv	g>a transition at 198 (= W66STOP)	a deletion at 269		a deletion at 269
Hv	t>c transition at 433 (= W145R)			a deletion at 269
Hvi	t>c transition at 148 (= C50R)			
Hvii	t>c transition at 148 (= C50R)			
Hviii	t>c transition at 148 (= C50R)			
Ii	c deletion at 391	g deletion at 583	a deletion at 269	t>g transversion at 433 (= W145G)
Iii	c deletion at 391	a deletion at 269	a deletion at 269	g>t transversion at 160 (= G54STOP)
Iiii	c deletion at 391	Inconclusive	a deletion at 269	t>g transversion at 433 (= W145G)
Iiv	c deletion at 391	g>a transition at 33 (= W11STOP)	c deletion at 390	
Iv	c deletion at 391	g insertion at 55	a deletion at 269	
Ivi		g>a transition at 52 (= G18S)	a deletion at 269	
Ji	g>t transversion at 334 (= E112STOP)	t>c transition at 433 (= W145R)	a deletion at 326	g deletion at 52
Jii	g>t transversion at 293 (= S98I)	g>a transition at 122 (= G41D)	a deletion at 326	gc deletion at 315
Jiii	g deletion at 502	g>a transition at 449 (= W150STOP)	a deletion at 326	gc deletion at 315
Jiv	g>a transition at 429 (= W143STOP)	t>c transition at 433 (= W145R)	Wild type	incomplete…resequence
Jv		t>c transition at 433 (= W145R)	g>a transition at 33 (= W11STOP)	incomplete…resequence
Jvi			g insertion at 133	
Ki	c deletion at 516	g deletion at 123	g>a transition at 50 (= W17STOP)	incomplete…resequence
Kii	c deletion at 516	Inconclusive	t>g transversion at 433 (= W145G)	g deletion at 122
Kiii	g deletion at 450	c>a transversion at 150 (= C50STOP)	a deletion at 326	incomplete…resequence
Kiv	g deletion at 450	g deletion at 200	a deletion at 326	incomplete…resequence
Kv		a deletion at 327	a deletion at 326	t>g transversion at 433 (= W145G)
Li	g>a transition at 33 (= W11STOP)	g insertion at 202	ag insertion at 120	gc deletion at 569
Lii	19 base deletion at 180	g>a transition at 434 (= W145 STOP)	g deletion at 450	g>t transversion at 151 (= D51Y)
Liii	c deletion at 228	g>a transition at 434 (= W145 STOP)		incomplete…resequence
Liv	c deletion at 228	g>a transition at 434 (= W145 STOP)		g>a transition at 51 (= W17STOP)
Lv	g>a transition at 33 (=W11STOP)	g>a transition at 434 (= W145 STOP)		incomplete…resequence
Lvi	c deletion at 228	g>a transition at 434 (= W145 STOP)		incomplete…resequence
Lvii	insertion at 176	g>a transition at 434 (= W145 STOP)		
Mi	20 base insertion at 180	a deletion at 269	33 base deletion at 262	incomplete…resequence

Mii	19 base insertion at 176	a deletion at 269	33 base deletion at 262	incomplete…resequence
Miii	19 base insertion at 176	a deletion at 46	33 base deletion at 262	incomplete…resequence
Miv	19 base insertion at 176	Inconclusive	incomplete…resequence	incomplete…resequence
Mv	19 base insertion at 176	g>t transversion at 79 (= E27STOP)	incomplete…resequence	g>a transition at 90 (= W30STOP)
Ni	g>t transversion at 124 (=G42C)	g>a transition at 295 (= D99N)	t>g transversion at 433 (= W145G)	g>a transition at 198 (= W66STOP)
Nii	inconclusive	g>a transition at 198 (= W66STOP)	t>g transversion at 433 (= W145G)	incomplete…resequence
Niii	t>c transition at 433 (= W145R)	g>a transition at 295 (= D99N)	incomplete…resequence	incomplete…resequence
Niv	t>c transition at 433 (= W145R)	g>a transition 8 nucleotides upstream from promoter		g>a transition at 198 (= W66STOP)
Oi	g>t transversion at 124 (=G42C)	g deletion at 161	t>c transition at 433 (= W145R)	g>t transversion at 412 (= G138STOP)
Oii	g>t transversion at 124 (=G42C)	g>a transition at 50 (= W17STOP)	t>c transition at 433 (= W145R)	incomplete…resequence
Oiii	g>t transversion at 124 (=G42C)	g>a transition at 50 (= W17STOP)	g>a transition at 50 (= W17STOP)	g>t transversion at 412 (= G138STOP)
Oiv	g>t transversion at 124 (=G42C)	g>a transition at 50 (= W17STOP)	t>c transition at 433 (= W145R)	incomplete…resequence
Ov	g>t transversion at 124 (=G42C)		g insertion at 133	
Pi	g>a transition at 198 (= W66STOP)	a deletion at 326	g>a transition at 125 (= G42D)	incomplete…resequence
Pii	g>a transition at 198 (= W66STOP)	a deletion at 326	incomplete…resequence	a deletion at 269
Piii	g>a transition at 198 (= W66STOP)	a deletion at 326	incomplete…resequence	a deletion at 269
Piv	g>a transition at 198 (= W66STOP)	a deletion at 326	t>c transition at 433 (= W145R)	a deletion at 269
Pv	g>a transition at 198 (= W66STOP)	a deletion at 326	incomplete…resequence	a deletion at 269
Pvi	g>a transition at 198 (= W66STOP)		t>c transition at 433 (= W145R)	a deletion at 269
Pvii	g>a transition at 460 (= E>K)		incomplete…resequence	
Pviii	inconclusive		incomplete…resequence	
Qi	g>a transition at 198 (= W66STOP)	g deletion at 200	incomplete…resequence	g deletion at 58
Qii	t>c transition at 433 (= W145R)	g deletion at 200	incomplete…resequence	a deletion at 269
Qiii	t>c transition at 433 (= W145R)	g deletion at 122		g>t transversion at 412 (= G138STOP)
Qiv	t>c transition at 433 (= W145R)	a deletion at 269		g deletion at 58
Qv	t>c transition at 433 (= W145R)	a deletion at 269		a deletion at 269
Qvi	t>c transition at 433 (= W145R)			a deletion at 269
Ri	g>t transversion at 124 (=G42C)	g>a transition at 51 (= W17STOP)	incomplete…resequence	g deletion at 86
Rii	g deletion at 430	t>c transition at 433 (= W145R)	incomplete…resequence	g deletion at 86
Riii		g>a transition at 51 (= W17STOP)	g>t transversion at 433 (= W145G)	g deletion at 86
Riv		3 base insertion at 127	incomplete…resequence	g deletion at 503
Rv			incomplete…resequence	g>t transversion at 124 (= G42C)
Rvi			incomplete…resequence	a>t transversion at 229 (= K77STOP)

Si	t>c transition at 433 (= W145R)	g deletion at 69	incomplete…resequence	incomplete…resequence
Sii	t>c transition at 433 (= W145R)	t>c transition at 433 (= W145R)	t>g transversion at 433 (= W145G)	
Siii	t>c transition at 433 (= W145R)	g>a transition at 429 (= W143STOP)	t>g transversion at 433 (= W145G)	
Siv	t>g transversion at 173 (= L58R)	a deletion at 326	t>g transversion at 433 (= W145G)	
Sv		g deletion at 69	t>g transversion at 433 (= W145G)	
Ti	t>c transition at 148 (= C50R)	incomplete sequence	g>a transition at 198 (= W66STOP)	a deletion at 269
Tii	t>c transition at 148 (= C50R)	g deletion at 69	g>a transition at 198 (= W66STOP)	a deletion at 269
Tiii	t>c transition at 148 (= C50R)	a>g transition at 464 (= D155G)		
Tiv	g>a transition at 295	g deletion at 123		
Tv	t>c transition at 148 (= C50R)	a deletion at 326		
Tvi		g>a transition at 33 (= W11STOP)		
Ui	g>a transition at 198 (= W66STOP)	g>a transition at 90 (= W30STOP)	g>a transition at 198 (= W66STOP)	g>a transition at 32 (= W11STOP)
Uii	g>a transition at 198 (= W66STOP)	g>a transition at 90 (= W30STOP)		c>t transition at 196 (= W66R)
Uiii	g>a transition at 122 (= G41D)	g>a transition at 90 (= W30STOP)		c>t transition at 196 (= W66R)
Vi	a insertion at 468	a deletion at 326	t>c transition at 433 (= W145R)	a>g transition at 80 (= E27G)
Vii		a deletion at 326	t>c transition at 433 (= W145R)	
Viii		a deletion at 326		
Viv				
Wi	g>a transition at 435 (= W145STOP)	a deletion at 326	g>a transition at 122 (= G41D)	incomplete…resequence
Wii	g>a transition at 435 (= W145STOP)	a deletion at 326	g>a transition at 122 (= G41D)	
Wiii	t>g transversion at 433 (= W145G)	a deletion at 500	g>a transition at 122 (= G41D)	
Wiv	a deletion at 269		g>a transition at 122 (= G41D)	
Xi	Incomplete…resequence	a deletion at 269	c deletion at 216	a insertion at 270
Xii	t deletion at 13	a deletion at 269	incomplete…resequence	t>a transversion at 273
Xiii	Incomplete…resequence			g>a transition at 335
Yi	g>a transition at 464 (= D155G)	a deletion at 269	incomplete…resequence	g>a transition at 198 (= W66STOP)
Yii	g>a transition at 464 (= D155G)	a deletion at 269	g>t transversion at 151 (= D51Y)	
Yiii	Incomplete…resequence	a deletion at 269	c deletion at 391	
Zi	g>a transition at 33 (= W11STOP)		incomplete…resequence	a deletion at 326
Zii	g>a transition at 198 (= W66STOP)		g>t transversion at 82 (= E28STOP)	
AAi	a deletion at 566	g deletion at 431	incomplete…resequence	c deletion at 388
AAii		a deletion at 269		
AAiii		a deletion at 269		

ABi ABii ABiii	19 base insertion at 176 g insertion at 35	a deletion at 269 a deletion at 269 c>a transversion at 150 (= C50STOP)	incomplete…resequence	g>t transversion at 253 (= R85Y)
ACi ACii	Wild type	g>a transition at 129 (= W43STOP)	incomplete…resequence	incomplete…resequence
ADi	c>a transversion at 240 (= Y80STOP)	g>c transversion at 90 (= W30C)		
AEi	g>a transition at 198 (= W66STOP)	a deletion at 500		
AFi AFii AFiii	c deletion at 391	incomplete…resequence a deletion at 269 a deletion at 269		
AGi	g>a transition at 51 (= W17STOP)	incomplete…resequence		
AHi	Incomplete…resequence	incomplete…resequence		
AIi	Incomplete…resequence	incomplete…resequence		
AJi	g deletion at 58	t>g transversion at 433 (= W145G)		
AKi		g>a transition at 51 (= W17STOP)		
ALi		t>c transversion at 433 (= W145R)		
AMi		a deletion at 269		
ANi		incomplete…resequence		
AOi		a deletion at 269		
APi		a deletion at 269		
AQi		incomplete…resequence		
ARi		incomplete…resequence		
ASi		g>a transition at 129 (= W43STOP)		

Key: Black = unique mutation, green = sister/daughter mutation, yellow = incomplete sequence, red = anomalous sequence